AN EVALUATION OF SUPERRESOLUTION METHODS FOR TACTICAL RADIO DIRECTION FINDING (U)

by

William J.L. Read

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
REPORT NO. 1091

Canada

October 1991
Ottawa

National Defence / Défense nationale

AN EVALUATION OF SUPERRESOLUTION METHODS FOR TACTICAL RADIO DIRECTION FINDING (U)

by

William J.L. Read
Communication Electronic Warfare Section
Electronic Warfare Division

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
REPORT NO. 1091

PCN
041LK

October 1991
Ottawa

ABSTRACT

This report evaluates superresolution direction finding methods for land tactical narrowband ~~VHF/UHF~~ purposes. The DF methods are described and evaluated in depth using a common theoretical framework beginning with classical methods and ending with modern day eigenanalysis based methods. Based on this analysis, superresolution methods can be described in terms of a five step procedure where each step has a unique purpose in the estimation process. Different methods vary according to the operations performed at each step. This is useful in analyzing and comparing the theoretical performance of various DF methods under the same conditions. The results from simulations are also included to support the theoretical evaluations.

RESUME

Ce rapport contient une évaluation des techniques de radiogoniométrie à superrésolution pour des signaux tactiques terrestres VHF/UHF à bande passante étroite. Les méthodes de radiogoniométrie sont décrites et évaluées en détail à partir d'une théorie unifiée valide pour les méthodes classiques et les méthodes modernes utilisant des valeurs propres. Il est possible, à partir de cette théorie, de décrire les méthodes de superrésolution à l'aide d'une procédure en cinq étapes où chaque étape a un but unique dans le processus d'évaluation. Les différentes méthodes se différencient par les opérations effectuées à chaque étape. Il s'agit d'une caractéristique utile pour l'analyse et la comparaison, dans les mêmes conditions, des performances théoriques de diverses méthodes de radiogoniométrie. Les évaluations théoriques présentées sont étayées par des simulations.

EXECUTIVE SUMMARY

A considerable amount of work on superresolution radio direction finding (DF) methods has been reported in open literature sources over the last twenty years. However, due to the variability in approaches, it is difficult to make definitive statements as to the relative performance and merits of the various methods described. In this report a common theoretical framework is used to describe the various methods. The theoretical discussion presented in this report is based on the analysis as first presented in the open literature as well as new analysis. The new analysis was required to ensure a unified approach in the theoretical discussion as well as provide a firm theoretical basis for techniques which have been taken for granted in the open literature, but rarely discussed.

Simulations are also presented to illustrate relative performance of the various DF methods using the same input data. Since this report was written in support of tactical VHF/UHF DF, the simulations deal predominantly with multipath (fully coherent) signals since multipath has been perceived to be the most serious source of error in current operational DF systems.

To keep this report to a manageable size, only representative DF methods based on classical spectral estimation techniques and those which have evolved from these techniques have been described. For example, this includes the Bartlett method, Linear Prediction, MUSIC, and other similar estimators, but does not include the phase interferometer, Esprit and Maximum Likelihood methods. Additionally, analysis has been restricted to uniform linear arrays and white Gaussian noise statistics. Despite these limitations, the discussion in this report is reflective of the main body of research in DF over the last twenty years and before this report was written.

DF methods can be described in terms of three filter models, namely, the moving average filter, the autoregressive filter, and the autoregressive moving average filter. In terms of accuracy and computational speed, methods based on the all pole or autoregressive filter model have shown the greatest promise and for this reason were the main focus of this report.

It is shown that the autoregressive filter based methods can be generalized as a five step procedure. These steps include:

1. Estimation of the autocorrelation matrix.

2. Division of the autocorrelation matrix into a signal and noise subspace.

3. Generation of an enhanced inverse autocorrelation matrix.

4. Estimation of the all pole filter coefficients.

5. Estimation of the signal bearings from the filter coefficients.

Each of these steps is the subject of separate sections in this report

In terms of the resultant DF accuracy, and given the assumptions and restrictions made in this report, the root-MUSIC and root-Minimum Norm methods represent two of the best methods for DF. These methods are far superior to classical techniques but are in

no way optimum - more research is still required in each of the steps listed previously. Additionally, in practice, many of the assumptions and restrictions used in this report are often untrue, so that other areas of research include:

1. Extension of techniques used for uniform linear arrays to arbitrary array geometries, or development of comparable techniques.

2. Sensor calibration.

3. Estimation of noise statistics in an unknown coloured noise environment

4. Modifications to the signal model for actual multipath environments and mixed coherent/noncoherent signal environments.

5. Faster algorithms for realtime implementation.

Solutions to the above problems will be required before the promises of superresolution methods can be fully realized.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

1.0 INTRODUCTION

A considerable amount of research and development has been carried out over the last twenty years in advanced spectral estimation methods which has led directly to the development of advanced radio direction finding (DF) methods. The methods of interest are those which are intended to operate against conventional narrowband VHF/UHF radio signals (i.e., CW, AM, FM, and SSB with bandwidths less than 200 kHz) in a land-based tactical environment. This report discusses these methods and their potential for improving the capabilities of current conventional tactical DF systems (discussed in reference [1-1]).

Each of the methods are discussed in terms of their development philosophy and their ability to deal with various error sources including noise, co-channel interference, and multipath. Although the subject of advanced DF is not limited to theory, a fundamental understanding of the capabilities of these methods is essential if a successful operational system is to be built.

1.1 SYSTEM OVERVIEW

For the sake of simplicity, the sensor systems discussed in this report are assumed to consist of a linear array of N vertical dipole antennas, all of which are connected to N separate perfectly matched (in both gain and phase) and like-numbered receiver channels as shown in Figure 1.1.

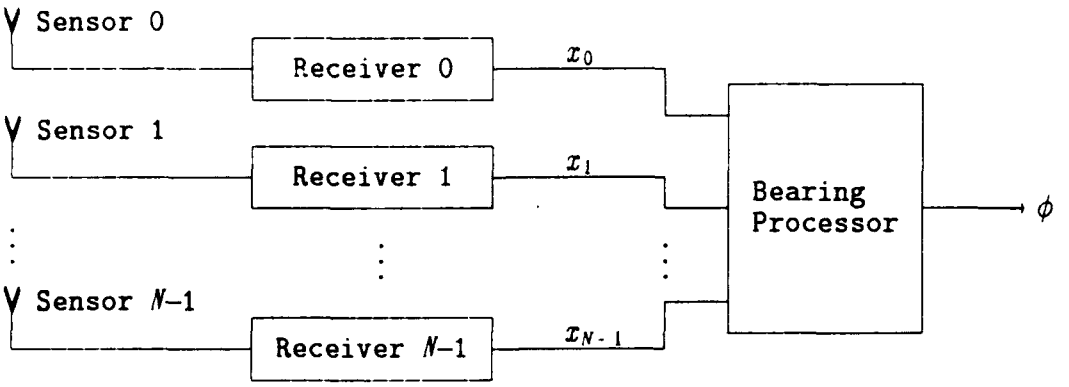


FIGURE 1.1: Block diagram of a multi-channel DF system

In the receiver, each channel is filtered to the correct bandwidth, mixed using a common reference signal (at the filter center frequency) to generate an in-phase and quadrature baseband representation of the antenna signal, which are then digitized. The sampling rate is assumed to be twice the bandwidth (i.e. the Nyquist rate). The bandwidth of the signal is assumed to be sufficiently narrow so that the modulation envelope does not change during the time taken for a signal to traverse the array when the signal is incident from boresight.

The receiver outputs are used as the inputs to the bearing processor which are represented by the complex valued parameters $x_0, x_1, x_2, \ldots, x_{N-1}$ where the subscript represents the channel number. The bearing processor inputs are assumed to be undistorted baseband representations of the antenna signals. In this context the values $x_0, x_1, x_2, \ldots, x_{N-1}$ are referred to as the sensor data. A single sample (alternately called a "snapshot" in the open literature) is defined as the set of values $x_0, x_1, x_2, \ldots, x_{N-1}$ measured at a specific instance in time. To distinguish sensor values from the same channel n but measured at different points in time (i.e. different time samples), an alternate representation is used, namely, $x_n(t)$, where t represents the sample index.

Throughout this report, the parameters M, N, and T are used exclusively to represent the number of signals, the number of sensors, and the number of sensor data samples, respectively. Other definitions and conventions followed throughout this report are described in the glossary.

1.2 SPECTRAL ESTIMATION AND RADIO DIRECTION FINDING

The application of spectral estimation concepts to radio direction finding is illustrated using Figure 1.2. In this representation a single radio signal in a noiseless environment traverses an antenna array with a bearing in azimuth given by ϕ. The assumption is made here (and throughout the rest of the report) that the transmitter is far enough from the antenna array so that the wavefront is planar. Additionally, since at VHF/UHF signal propagation is generally limited to ground waves, elevation angles are always considered to be 0 degrees.

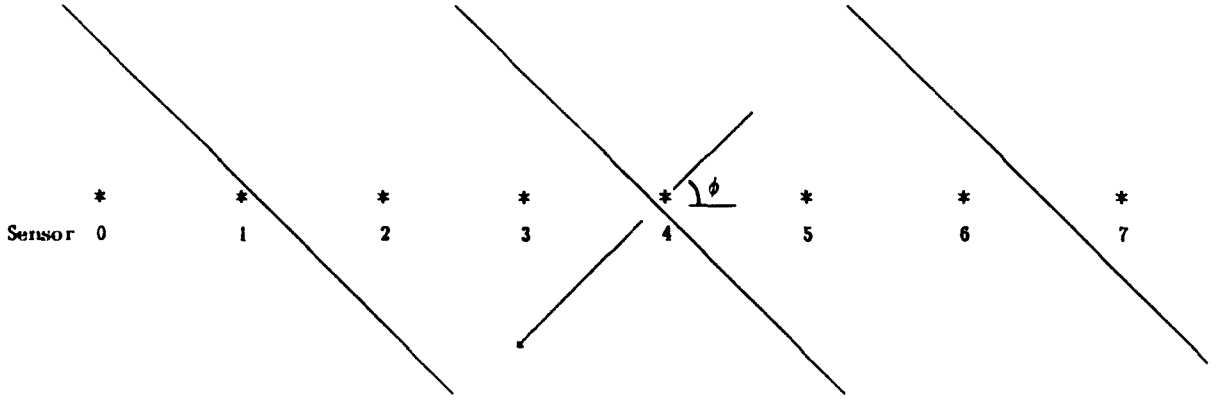


FIGURE 1.2: Radio direction finding sensor array

The amplitude and phase measured at sensor n in Figure 1.2 is represented at baseband by the complex value,

$$x_n = c(t)e^{-jn\omega d}, \tag{1.1}$$

where $c(t)$ is the complex baseband modulation envelope measured at sensor 0, d is the

sensor spacing, and ω is the *spatial* frequency defined by

$$\omega = \frac{2\pi}{\lambda}\cos(\phi) \ \text{rad/m}. \tag{1.2}$$

Here λ represents the signal wavelength. The term spatial frequency is used to denote that the frequency is measured with respect to position rather than with respect to time. For more complex antenna configurations, the spatial frequency may still be calculated using equation (1.2) by designating an arbitrary baseline as the reference baseline and then making all angular measurements with respect to it.

If a number of signals of different bearings impinge upon the sensor system in a noisy environment, then the spatial signal x_n is the sum of the individual signals, that is,

$$x_n = \sum_{m=1}^{M} c_m(t) e^{-jn\omega_m d} + \eta_n(t). \tag{1.3}$$

where M represents the total number of signals and $\eta_n(t)$ is the noise measured at sensor n. Spectral estimation methods can be used to resolve the various components in equation (1.3) by first generating the spatial power spectral density function. By taking advantage of the relationship between spatial frequency and bearing, this spectrum is converted to a DF spectrum which gives the power density versus bearing. The location of the peaks (local maximums) in the DF spectrum are then used as estimates for the signal bearings.

The spectral estimation approach has led to the development of a number sophisticated direction finding methods whose origins are based on research done in fields as diverse as geophysics, digital communications, underwater acoustics, and speech analysis. In many of these applications spectral estimation methods are used to process time series (or temporal) data, whereas in the direction finding case these methods are used to process position series (or spatial) data.

Mathematically there are some important differences between temporal frequency estimation methods and spatial frequency estimation methods. One difference is that position is a three dimensional parameter while time is one dimensional. Unless otherwise indicated, the DF methods in this report are discussed in terms of a single antenna baseline with uniform antenna spacing. In this special case, position is restricted to one dimension and the correspondence between the temporal and spatial methods is one to one. In general, results for single baselines can be extended to multiple baselines, if care is used.

Another important difference between temporal and spatial methods is that in spatial frequency estimation the time dimension has no equivalence in temporal frequency estimation. This extra dimension provides information which is useful when dealing with uncorrelated signals or temporal noise. It is also easier and a lot less expensive to sample with respect to time than with respect to position.

The layout of this report, where applicable, follows the historical development of the DF methods discussed, starting with classical methods (Section 2), followed by model based methods (Sections 4-5), then on to the more advanced eigenanalysis methods (Sections 6-8), and finally ending with a discussion of the rooting method (Section 9). Additional sections have also been included for the discussion of methods which can be considered separately from DF, but are nonetheless critical to the success of advanced DF

methods. This includes autocorrelation matrix estimation (Section 3), and model order determination (Section 9).

The methods discussed in this report are, for the most part, based on the autocorrelation matrix formulations since this simplifies comparisons. In many cases variants of these methods exist which are computationally more attractive, but are not discussed since they are theoretically identical.

2.0 THE CLASSICAL APPROACH

The exact DF spectrum can be calculated by computing the spatial autocorrelation sequence and then taking the Fourier transform of this sequence. This is called the indirect method. Alternatively, the exact DF spectrum can be calculated by taking the Fourier transform of the input data sequence and squaring the absolute values of the transformed sequence. This is called the direct method.

In either the direct or indirect methods, the assumption is made that there is an infinite amount of data available. Under realistic conditions, this clearly will not be the case, so instead approximations must be used which result in DF spectrums which are only estimations of the true function. Additionally, under these conditions, the two methods may produce different results.

2.1 THE INDIRECT METHOD OF ESTIMATING THE DF SPECTRUM

The DF spectrum can be defined as the Fourier transform of the spatial autocorrelation sequence and is given by,

$$S(\phi) = \sum_{m=-\infty}^{\infty} r_{xx}(m)\, e^{-j\omega md} \tag{2.1}$$

where the spatial frequency ω is related to the bearing angle ϕ by equation (1.2), and d is the spacing between consecutive antennas. The definition and estimation of the autocorrelation sequence $r_{xx}(m)$ is discussed in the next two sections.

2.1.1 The Autocorrelation Sequence

The statistical definition of the autocorrelation of a spatially sampled random process, x_k, at two different indices m and n is defined here as

$$r_{xx}(m,n) = E\{x_m\, x_n^*\} \tag{2.2}$$

where $E\{y\}$ is the mean or expected value of the process represented by y. If the sampled process is wide sense stationary, then its mean is constant for all indices, or

$$E\{x_m\} = E\{x_n\}, \tag{2.3}$$

and its autocorrelation will depend only on the difference n-m, or

$$E\{x_m\, x_n^*\} = E\{x_{m+k}\, x_{n+k}^*\}, \tag{2.4}$$

where k is any arbitrary integer. Under these conditions, an autocorrelation sequence may be defined where each element is given by

$$r_{xx}(m) = E\{x_{n+m}\, x_n^*\}. \tag{2.5}$$

The index value m is called the autocorrelation lag.

In direction finding, the wide sense stationary condition means that all signal

sources must be located in the farfield of the antenna array (i.e. the wavefront arriving at the antenna array produced by any one source is planar) and the first and second order statistics of the noise (mean and variance) do not change with respect to position and time.

2.1.2 The Unbiased Estimator

A more practical definition of the exact autocorrelation sequence than the one defined by equation (2.5), which takes advantage of the fact that the input data is sampled in both position and time, is given by

$$r_{xx}(m) = \lim_{T,N \to \infty} \frac{1}{2T+1} \sum_{t=-T}^{T} \frac{1}{2N+1} \sum_{n=-N}^{N} x_{n+m}(t)\, x_n^*(t), \tag{2.6}$$

where T represents the number of time samples and N represents the number of sensors (these definitions for T and N are used throughout the rest of this report). Here t represents the temporal index and n the spatial index. The Fourier transform of this sequence is the exact spectrum. Unfortunately, under realistic conditions, the number of data values is finite and consequently the value of N is also finite, rather than infinite. As a result, it will only be possible to estimate the autocorrelation sequence.

One method of estimating the sequence is to modify equation (2.6) to give,

$$\hat{r}_{xx}(m) = \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N-m} \sum_{n=0}^{N-m-1} x_{n+m}(t)\, x_n^*(t) \qquad \text{for } 0 \le m < N, \tag{2.7}$$

$$\hat{r}_{xx}(m) = 0 \qquad \text{for } m \ge N, \tag{2.8}$$

and,

$$\hat{r}_{xx}(m) = \hat{r}_{xx}^*(-m) \qquad \text{for } m < 0. \tag{2.9}$$

Equation (2.7) is an unbiased estimate of the autocorrelation sequence, since the expected value of the estimated autocorrelation lag is the true value.

Since only estimates of the autocorrelation sequence are available, the Fourier transform of this sequence will be an estimate of the DF spectrum. That is,

$$\hat{S}(\phi) = \sum_{m=1-N}^{N-1} \hat{r}_{xx}(m)\, e^{-j\omega m d}, \tag{2.10}$$

where the values of $\hat{r}_{xx}(m)$ are computed using equations (2.7) and (2.9).

2.1.3 The Biased Estimator

One difficulty with generating the autocorrelation sequence using equation (2.7), especially if only a few samples are available, is that the variances of the estimated autocorrelation sequence values increase as the absolute lag index increases, since there are

less and less data values available for averaging. As a consequence of this, it is possible for the estimated sequence to be non-realizable. For example, the estimate of any value in the autocorrelation sequence (except at lag 0) may result in a value larger than the one at lag 0, which is not physically possible.

To overcome this problem, a triangular weighting function (window) is applied which results in the modified estimate given by

$$\hat{r}_{xx}(m) = \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=0}^{N-m-1} x_{n+m}(t)\, x_n^*(t). \tag{2.11}$$

This is a biased estimate of the autocorrelation sequence, since the expected value of estimated autocorrelation lag is not the true value, except for lag 0. However, this estimate always results in a physically realizable autocorrelation sequence. Additionally, as the number of values N approaches infinity, the estimated autocorrelation sequence approaches the true sequence. In this context, the estimator is asymptotically unbiased.

A second advantage of the biased autocorrelation sequence estimate, compared to the unbiased estimate, is a lower variance since the effect of the larger lag elements, which are estimated from fewer product terms and therefore have greater variance, are suppressed. The disadvantage is that windowing degrades the resolution of the corresponding DF spectrum estimate (which is computed using equations (2.10) and (2.11))

2.2 THE DIRECT METHOD OF ESTIMATING THE DF SPECTRUM

An alternate method (called the direct method here) of computing the DF spectrum is to square the absolute magnitude of the Fourier transform calculated directly from the data, that is,

$$S(\phi) = \frac{1}{N} |X(\phi)|^2 = \frac{1}{N} X(\phi)\, X(\phi)^*, \tag{2.12}$$

where $X(\phi)$ represents the spatial Fourier transform of the data.

If only a finite amount of data is available, $X(\phi)$ can be approximated by

$$\hat{X}(\phi) = \sum_{m=0}^{N-1} x_m(t)\, e^{-j\omega md} \tag{2.13}$$

where d is the antenna spacing. Using this approximation, the calculated value of $S(\phi)$ will be an estimate of the true value. Substituting equation (2.13) into (2.12) gives

$$\hat{S}(\phi) = \frac{1}{N} \left(\sum_{m=0}^{N-1} x_m(t)\, e^{-j\omega md} \right) \left(\sum_{m=0}^{N-1} x_m^*(t)\, e^{+j\omega md} \right). \tag{2.14}$$

The spectrum generated for a random process using equation (2.14) is statistically unstable [2-1]. That is, as N increases, the variance of the estimated value of $S(\phi)$ does not approach zero, but rather becomes proportional to the square of the true value of $S(\phi)$. The method used to get around this difficulty is to compute the estimated spectrum for

several different time samples, then average the spectrums together to give an average version. That is,

$$\hat{S}(\phi) = \frac{1}{NT} \sum_{t=0}^{T-1} \left[\left(\sum_{m=0}^{N-1} x_m(t)\, e^{-j\omega md} \right) \left(\sum_{m=0}^{N-1} x_m^*(t)\, e^{+j\omega md} \right) \right]. \tag{2.15}$$

2.3 MATRIX REPRESENTATION OF THE CLASSICAL ESTIMATORS

2.3.1 The Direct Method

Inspection of equation (2.13), the Fourier transform of the sensor data, suggests a compact matrix representation which is,

$$\hat{X}(\phi) = \mathbf{e}^H \mathbf{x}_t, \tag{2.16}$$

where $\mathbf{e}$ and $\mathbf{x}_t$ are both N element column vectors. The vector $\mathbf{x}_t$ is the sensor data vector measured at a time t, and is defined as

$$\mathbf{x}_t = \begin{bmatrix} x_0(t) \\ x_1(t) \\ x_2(t) \\ \vdots \\ x_{N-1}(t) \end{bmatrix}. \tag{2.17}$$

The N element vector $\mathbf{e}$ is referred to as the steering vector and is given by

$$\mathbf{e} = \begin{bmatrix} 1 \\ e^{+j\omega d} \\ e^{+j\omega 2d} \\ \vdots \\ e^{+j\omega(N-1)d} \end{bmatrix}. \tag{2.18}$$

The estimated DF spectrum, using the direct method (equation (2.12)), is then

$$\hat{S}(\phi) = \frac{1}{N} \mathbf{e}^H \mathbf{x}_t \mathbf{x}_t^H \mathbf{e}. \tag{2.19}$$

Following the example of equation (2.13) and incorporating time averaging to generate a more stable estimate gives,

$$\hat{S}(\phi) = \frac{1}{NT} \sum_{t=0}^{T-1} \mathbf{e}^H \mathbf{x}_t \mathbf{x}_t^H \mathbf{e}, \tag{2.20}$$

or,

$$\hat{S}(\phi) = \frac{1}{N} \mathbf{e}^{\mathrm{H}} \mathbf{X} \mathbf{X}^{\mathrm{H}} \mathbf{e}, \tag{2.21}$$

where

$$\mathbf{X} = \frac{1}{\sqrt{T}} \begin{bmatrix} x_0(0), & x_0(1), & x_0(2), \ldots, & x_0(T\text{-}1) \\ x_1(0), & x_1(1), & x_1(2), \ldots, & x_1(T\text{-}1) \\ x_2(0), & x_2(1), & x_2(2), \ldots, & x_2(T\text{-}1) \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-1}(0), & x_{N-1}(1), & x_{N-1}(2), \ldots, & x_{N-1}(T\text{-}1) \end{bmatrix}. \tag{2.22}$$

Equation (2.21) is called the Bartlett estimator [2-2] (note that since for bearing estimation purposes only the shape of the spectrum is important the factor $1/N$ can be dropped from this expression).

2.3.2 The Indirect Method

An alternate representation of the Bartlett estimator given by equation (2.21) can be formed be letting the matrix product $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ be represented by a single matrix so that,

$$\hat{S}(\phi) = \frac{1}{N} \mathbf{e}^{\mathrm{H}} \hat{\mathbf{R}} \mathbf{e}. \tag{2.23}$$

The elements of the matrix $\hat{\mathbf{R}}$ are given by

$$\hat{r}_{ij} = \frac{1}{NT} \sum_{t=0}^{T-1} x_i(t)\, x_j^*(t). \tag{2.24}$$

The maximum stability in the estimate of the spectrum occurs as the number of samples, T, is increased to infinity. In this limiting case the elements of the matrix become,

$$\mathrm{E}\{\hat{r}_{ij}\} = \mathrm{E}\{x_i\, x_j^*\} = r_{xx}(i\text{-}j), \tag{2.25}$$

which are the values of the spatial autocorrelation sequence. As a result, $\mathbf{R}$ is called the autocorrelation matrix and can be defined in terms of the estimated autocorrelation matrix as,

$$\mathbf{R} = \mathrm{E}\{\hat{\mathbf{R}}\}. \tag{2.26}$$

The matrix $\mathbf{R}$ is also called the covariance matrix since r_{ij} also represents the covariance between two time varying processes defined by $x_i(t)$ and $x_j(t)$.

If the matrix calculations are performed for equation (2.23) the result can be expressed as,

$$\hat{S}(\phi) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} e^{-j\omega md} \hat{r}_{mn} e^{+j\omega nd}. \tag{2.27}$$

Replacing $\hat{r}_{mn}$ with equation (2.24) gives

$$\hat{S}(\phi) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \frac{1}{T} \sum_{t=0}^{T-1} x_m(t)\, x_n^*(t)\, e^{+j\omega(n-m)d}. \tag{2.28}$$

Grouping terms with the same Fourier coefficient $e^{+j\omega(m-n)d}$ and expanding,

$$\begin{aligned}\hat{S}(\phi) = \frac{1}{NT} [&(x_0(t)\, x_{N-1}^*(t))\, e^{+j\omega(N-1)d} + (x_0(t)\, x_{N-2}^*(t) + x_1(t)\, x_{N-1}^*(t)) e^{+j\omega(N-2)d} \\ &+ (x_0(t)\, x_{N-3}^*(t) + x_1(t)\, x_{N-2}^*(t) + x_2(t)\, x_{N-1}^*(t)) e^{+j\omega(N-3)d} \\ &+ \ldots + (x_{N-1}(t)\, x_0^*(t))\, e^{-j\omega(N-1)d}],\end{aligned} \tag{2.29}$$

and rewriting,

$$\hat{S}(\phi) = \sum_{m=1-N}^{N-1} e^{+j\omega md} \left[\frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=0}^{N-m-1} x_{n+m}(t)\, x_n^*(t) \right]. \tag{2.30}$$

Inspection of the quantity in the square brackets above reveals it to be the biased autocorrelation sequence estimator defined earlier by equation (2.11). From this, it is apparent that the use of the autocorrelation matrix in equation (2.23) is equivalent to the direct method if, and only if, the biased autocorrelation sequence estimator is used.

2.4 DATA WINDOWING

There are several consequences of the spatially limited data available. For one, the spatial frequency resolution of this technique is limited to roughly the reciprocal of the antenna baseline, or worse. Another consequence is that the sudden transition to zero of the unknown values in the data sequence has the effect of causing large side lobes (Gibbs phenomenon) in the spatial frequency spectrum. Window functions can be applied which taper the data at the start and end of the data sequence to reduce the sidelobe problem. As a consequence, the data is modified to become

$$y_n = a_n\, x_n \tag{2.31}$$

where a_n is the weighting coefficient. The modified value y_n is used in place of x_n in any of the estimation techniques discussed so far.

In terms of the matrix equations discussed previously, the modification given in equation (2.31) results in a new data matrix **Y** (which replaces **X**) of the form,

$$\mathbf{Y} = \frac{1}{\sqrt{T}} \begin{bmatrix} a_0x_0(0), & a_0x_0(1), & a_0x_0(2), \ldots, & a_0x_0(T\text{-}1) \\ a_1x_1(0), & a_1x_1(1), & a_1x_1(2), \ldots, & a_1x_1(T\text{-}1) \\ a_2x_2(0), & a_2x_2(1), & a_2x_2(2), \ldots, & a_2x_2(T\text{-}1) \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-1}x_{N-1}(0), & a_{N-1}x_{N-1}(1), & a_{N-1}x_{N-1}(2), \ldots, & a_{N-1}x_{N-1}(T\text{-}1) \end{bmatrix}. \tag{2.32}$$

The effects of using different types of data windows are given in reference [2-3]. The application of the weighting function has a price, however, as sidelobe reduction is achieved at the expense of resolution.

2.5 EXTENSION TO NONLINEAR AND NONUNIFORM ARRAYS

Inspection of the steering vector, $\mathbf{e}$, reveals that it contains the (conjugate) spatial Fourier transform coefficients corresponding to each antenna in the array. In spatial terms, the steering vector represents a sinusoidal signal propagating through space in a known direction. Each element in the vector defines the amplitude and phase parameters of the signal at each antenna position relative to antenna position 0.

For a sinusoidal wave propagating in a known direction, the steering vector elements can be defined for any antenna position as

$$e_n = e^{+j\frac{2\pi}{\lambda}(x_n\cos(\phi)\cos(\psi) + y_n\sin(\phi)\cos(\psi) + z_n\sin(\psi))} \tag{2.33}$$

where the n^{th} antenna position is specified in terms of the cartesian coordinates (x_n, y_n, z_n), antenna 0 is located at the origin, the x-y plane is the surface of a flat Earth, and e_n is the n^{th} element of the steering vector. Additionally the elevation bearing angle ψ has also been added for the more general case where the signal elevation angle must also be determined (this changes the bearing estimation procedure from a one dimensional search in azimuth to a two dimensional search in both azimuth and elevation).

Equation (2.33) may be written in a more compact form, namely,

$$e_n = e^{+j\frac{2\pi}{\lambda}(\mathbf{P}_n^{\mathrm{T}}\mathbf{B})}, \tag{2.34}$$

where $\mathbf{P}_n$ is the vector representing the position of the n^{th} antenna and is defined in Cartesian coordinates as

$$\mathbf{P}_n = \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix}, \tag{2.35}$$

and $\mathbf{B}$ is the signal direction vector defined by

$$\mathbf{B} = \begin{bmatrix} \cos(\phi)\cos(\psi) \\ \sin(\phi)\sin(\psi) \\ \sin(\psi) \end{bmatrix}. \tag{2.36}$$

For a uniform linear array, where the array baseline is aligned with the x axis (i.e. $\mathbf{P}_n = [nd, 0, 0]^T$) equation (2.34) reduces to a more familiar form (see equation (2.18)) where,

$$e_n = e^{+j\omega nd}. \tag{2.37}$$

2.6 BEAMFORMING

The directional response of the spectral estimators described to this point are equivalent to beamforming [2-4]. For example, the general form for the output of a beamformer is given by

$$y(t) = \sum_{m=0}^{N-1} a_m \, x_m(t - t_m), \tag{2.38}$$

where a_m represents the weighting coefficients and t_m the time delays. In this example $x_m(t)$ represents the carrier modulated signal (not the baseband representation used throughout the rest of this report) which is given by,

$$x_m(t) = c_m(t) e^{j\omega r_m} e^{j\beta t} \tag{2.39}$$

where $c_m(t)$ represents the complex amplitude of the signal measured at sensor m, ω represents the spatial frequency, β represents the carrier frequency, and r_m represents the displacement from sensor 0 in a direction parallel to the reference baseline (note that $r_m = md$ for a uniform linear array).

If the values of the time delays t_m are small (i.e. $t_m << 0.5/BW$ where BW represents the bandwidth of the modulating signal $c_m(t)$) then the beamformer output can be accurately approximated by,

$$y(t) = \sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{-j\beta t_m}. \tag{2.40}$$

Additionally, by choosing the time delays so that,

$$\beta t_m = \omega r_m, \tag{2.41}$$

the beamformer equation becomes,

$$y(t) = \sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{-j\omega r_m}. \tag{2.42}$$

Inspection of this last result shows that for appropriate choice of time delays, the

beamformer output $y(t)$ is the spatial Fourier transform of the weighted sensor data at time instance t (similar to equation (2.13)). That is, the beamformer is an alternate implementation of the classical methods described previously.

Following the same procedure as before to compute the DF spectrum (see section 2.2), then

$$\hat{S}(\phi) = \frac{1}{NT} \sum_{t=0}^{T-1} \left[\left(\sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{-j\omega r_m} \right) \left(\sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{+j\omega r_m} \right) \right], \tag{2.43}$$

or in the more general case,

$$\hat{S}(\phi) = \frac{1}{NT} \sum_{t=0}^{T-1} \left[\left(\sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{-j\beta t_m} \right) \left(\sum_{m=0}^{N-1} a_m \, x_m(t) \, e^{+j\beta t_m} \right) \right]. \tag{2.44}$$

Expressed in matrix form

$$\hat{S}(\phi) = \frac{1}{N} \mathbf{g}^{\mathrm{H}} \mathbf{Y} \mathbf{Y}^{\mathrm{H}} \mathbf{g}, \tag{2.45}$$

where the vector $\mathbf{g}$ is defined as

$$\mathbf{g} = \begin{bmatrix} e^{+j\beta t_0} \\ e^{+j\beta t_1} \\ e^{+j\beta t_2} \\ \vdots \\ e^{+j\beta t_{N-1}} \end{bmatrix}, \tag{2.46}$$

and the matrix $\mathbf{Y}$ was defined in section 2.4.

Equation (2.44) is identical in form to the classical spectral estimator described by equation (2.21) except that the windowed data matrix $\mathbf{Y}$ is used in place of the matrix $\mathbf{X}$, and the vector $\mathbf{g}$ is a more general version of the steering vector $\mathbf{e}$. The two vectors are identical if equation (2.41) is satisfied or the elements of the vector $\mathbf{g}$ are chosen so that

$$g_n = e_n = e^{+j\frac{2\pi}{\lambda}(\mathbf{P}_n{}^{\mathrm{T}}\mathbf{B})}, \tag{2.47}$$

where $\mathbf{P}_n$ and $\mathbf{B}$ were described in section 2.5. In terms of the time delay coefficients used in the beamforming network, this is equivalent to,

$$t_m = \frac{\mathbf{P}_n{}^{\mathrm{T}}\mathbf{B}}{c}, \tag{2.48}$$

where c is the propagation speed of light.

3.0 THE ESTIMATED AUTOCORRELATION MATRIX

The autocorrelation matrix, which was first introduced in conjunction with classical direction finding algorithms, is a central feature in the derivation of the superresolution algorithms to follow. All these algorithms are derived assuming the true autocorrelation is known, which is rarely true in practice. The autocorrelation matrix is normally estimated from the available input data, and the manner in which this is done can impact on the performance of the superresolution algorithm especially in the case of a small number of sensors. Given the importance of the estimation procedure, some of the more common methods are discussed here along with some of their drawbacks.

Estimation of the autocorrelation matrix using only a single data sample will be discussed first since it simplifies the task of illustrating some of the concepts involved. Estimation involving several data samples measured at different instances in time is introduced later, followed by a discussion of the effects of noise.

Before proceeding with a description of the autocorrelation matrix estimation methods, it is useful to examine the structure of the true autocorrelation matrix as this will provide a basis for comparing various estimation methods. The definition of the autocorrelation matrix is given by,

$$\mathbf{R} = \mathrm{E}\{\mathbf{X}\mathbf{X}^{H}\}, \tag{3.1}$$

where $\mathbf{X}$ is a data matrix (or vector) and is described in the following sections. The dimensions of $\mathbf{R}$ are defined here to be $(p+1)\times(p+1)$ where the value of p is less than the number of sensors N. The definition of p has been chosen this way to correspond to the filter order of the all pole filter methods discussed later on in this report. If we assume that the received signal is made up of M signals with unique bearings plus uncorrelated sensor noise (either spatial and/or temporal), then the data matrix has the form

$$\mathbf{X} = \sum_{i=1}^{M} \mathbf{S}_i + \mathbf{N}. \tag{3.2}$$

where $\mathbf{S}_i$ is the data matrix formed from the i^{th} signal in the absence of all other signals or noise. Noting that the noise is uncorrelated and that signals arriving from different directions are also uncorrelated (with respect to position), then the autocorrelation matrix can be redefined as

$$\mathbf{R} = \mathrm{E}\{\sum_{i=1}^{M} \mathbf{S}_i\mathbf{S}_i^{H}\} + \mathrm{E}\{\mathbf{N}\mathbf{N}^{H}\}. \tag{3.3}$$

Since the signal is assumed to be deterministic (not random) the signal correlation matrix is defined in this report as,

$$\mathbf{R}_s = \sum_{i=1}^{M} \mathbf{S}_i\mathbf{S}_i^{H}, \tag{3.4}$$

and the noise correlation matrix as

$$\mathbf{R}_N = \mathrm{E}\{\mathbf{N}\mathbf{N}^{\mathrm{H}}\}. \tag{3.5}$$

An important aspect of the true autocorrelation matrix **R** is the rank of the signal correlation matrix $\mathbf{R}_s$. If the number of signals is less than the number of rows or columns in **R** (i.e. $M \leq p$) then $\mathbf{R}_s$ will have rank M. Based on this, it would be possible to set up M equations from the rows or columns of $\mathbf{R}_s$ to solve for the bearings of the M unknown signals exactly. This is the basis for improved performance of superresolution spectral estimators compared to classical estimators which are based on correlating an ideal signal with the data which requires substantially more data.

In practical situations, the true autocorrelation matrix **R** is not usually available and must be estimated from the data. In this case the bearings cannot be solved exactly, but must also be estimated generally using least mean square techniques. The accuracy of the bearings then becomes a function of the accuracy of the autocorrelation matrix estimate.

3.1 ESTIMATING THE AUTOCORRELATION MATRIX

The estimated autocorrelation matrix is formed from the input data using the following expression:

$$\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^{\mathrm{H}}. \tag{3.6}$$

The various methods of estimating the autocorrelation matrix discussed in the following sections differ only in the manner of setting up the data matrix **X**.

In the simplest case, the data matrix is a vector defined as

$$\mathbf{x}_f = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix}. \tag{3.7}$$

In this case the lower case form is used to denote a vector and the subscript f has been added to denote the forward data case (i.e., $\mathbf{X} = \mathbf{x}_f$). In the forward case, the data is ordered from 0 to N-1. In a similar manner, a backward data vector may also be defined as,

$$\mathbf{x}_b = \begin{bmatrix} * \\ x_{N-1} \\ * \\ x_{N-2} \\ * \\ x_{N-3} \\ \vdots \\ * \\ x_0 \end{bmatrix}. \tag{3.8}$$

The backward formulation follows from the form of the backwards signal model. For example, letting the forward data vector elements, x_{fn}, be represented by equation (1.3), the backward elements can be represented as,

$$x_{bn} = \sum_{m=1}^{M} c_m(t) e^{+\jmath((N-n)\omega_m d+\theta_m)} = \sum_{m=1}^{M} c_m(t) e^{-\jmath(n\omega_m d+\Omega_m)}. \quad (3.9)$$

In comparing the forward and backward signal models, the only difference in the two expressions are the signal phases represented by θ_m and Ω_m respectively. Since the apparent signal bearings are unchanged, the bearings determined using either data vector will be identical. The above analysis does not follow for noise, so that in the noisy case the bearings estimated using either data vector will only be approximately the same.

One difficulty in using either $\mathbf{x}_f$ or $\mathbf{x}_b$ to estimate the autocorrelation matrix using equation (3.6) is that since only one unique vector is used to make the estimate, the resulting matrix will have rank 1. Superresolution algorithms depend on the rank of the signal correlation matrix equaling the number of signals, so if the number of signals is greater than 1, resolution is severely degraded.

A second difficulty with computing the estimated autocorrelation matrix in this manner is that the elements of the matrix are sensitive to instantaneous noise perturbations in the data, whereas the true autocorrelation matrix is only sensitive to the the statistics of the noise which are assumed to be constant or very slowly changing over the measurement period.

Methods to deal with these difficulties are discussed in the following sections.

3.2 THE AUTOCORRELATION METHOD

One method, called the *autocorrelation* method [3-1], is to append p zeros to the start and p zeros to the end of the data. Variants include appending zeros to the start or end only. The data is then constructed from a number of shifted subarrays of size $p+1$ where $p < N$ as shown below,

$$\mathbf{X}_f = \frac{1}{\sqrt{N-p}} \begin{bmatrix} 0, & 0, & 0, & \ldots, & 0, & x_0, & \ldots, & x_{N-p-1}, & x_{N-p}, & \ldots, & x_{N-3}, & x_{N-2}, & x_{N-1} \\ 0, & 0, & 0, & \ldots, & x_0, & x_1, & \ldots, & x_{N-p}, & x_{N-p+1}, & \ldots, & x_{N-2}, & x_{N-1}, & 0 \\ 0, & 0, & 0, & \ldots, & x_1, & x_2, & \ldots, & x_{N-p+1}, & x_{N-p+2}, & \ldots, & x_{N-1}, & 0, & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0, & 0, & x_0, & \ldots, & x_{p-3}, & x_{p-2}, & \ldots, & x_{N-3}, & x_{N-2}, & \ldots, & 0, & 0, & 0 \\ 0, & x_0, & x_1, & \ldots, & x_{p-2}, & x_{p-1}, & \ldots, & x_{N-2}, & x_{N-1}, & \ldots, & 0, & 0, & 0 \\ x_0, & x_1, & x_2, & \ldots, & x_{p-1}, & x_p, & \ldots, & x_{N-1}, & 0, & \ldots, & 0, & 0, & 0 \end{bmatrix}, \quad (3.10)$$

where the subscript f is again used to denote the forward case. The corresponding conjugate backward matrix is given by

$$\mathbf{X}_b^* = \frac{1}{\sqrt{N-p}} \begin{bmatrix} 0, & 0, & 0, & \ldots, & 0, & x_{N-1}, & \ldots, & x_p, & x_{p-1}, & \ldots, & x_2, & x_1, & x_0 \\ 0, & 0, & 0, & \ldots, & x_{N-1}, & x_{N-2}, & \ldots, & x_{p-1}, & x_{p-2}, & \ldots, & x_1, & x_0, & 0 \\ 0, & 0, & 0, & \ldots, & x_{N-2}, & x_{N-3}, & \ldots, & x_{p-2}, & x_{p-3}, & \ldots, & x_0, & 0, & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0, & 0, & x_{N-1}, & \ldots, & x_{N-p+2}, & x_{N-p+1}, & \ldots, & x_2, & x_1, & \ldots, & 0, & 0, & 0 \\ 0, & x_{N-1}, & x_{N-2}, & \ldots, & x_{N-p+1}, & x_{N-p}, & \ldots, & x_1, & x_0, & \ldots, & 0, & 0, & 0 \\ x_{N-1}, & x_{N-2}, & x_{N-3}, & \ldots, & x_{N-p}, & x_{N-p-1}, & \ldots, & x_0, & 0, & \ldots, & 0, & 0, & 0 \end{bmatrix}. \quad (3.11)$$

Although the autocorrelation method avoids the rank deficiency problem discussed in the previous section, resolution of spectral estimates (including classical methods) are degraded due to the pre- and postwindowing of the data (i.e. the values of the unknown data in the data matrix $x_1, x_2, x_3, \ldots, x_p$ and $x_N, x_{N+1}, x_{N+2}, \ldots, x_{N+p-1}$ are assumed to be zero). This problem becomes more severe as the number of sensors is decreased. Figure 3.1 illustrates one example of this where three equal power signals with bearings of 40, 50, and 120 degrees are intercepted by an 8 element sensor array with one half wavelength spacing. The signal to noise ratio was 65 dB. Three methods (the autocorrelation method and two others to be described in the following sections) were used to estimate the autocorrelation matrix for $p = 4$. The Thermal Noise estimator (section 5.2.6) was then used to compute the corresponding DF spectrums. In the case of the autocorrelation method, the signals at 40 and 50 degrees were not resolved despite the high signal to noise ratio. In general, the results were poor compared with the other autocorrelation matrix estimation methods.

Based on the poor resolution of this method, the autocorrelation method is considered too inaccurate for systems with small arrays (e.g. tactical systems) and is not discussed in the rest of this report.

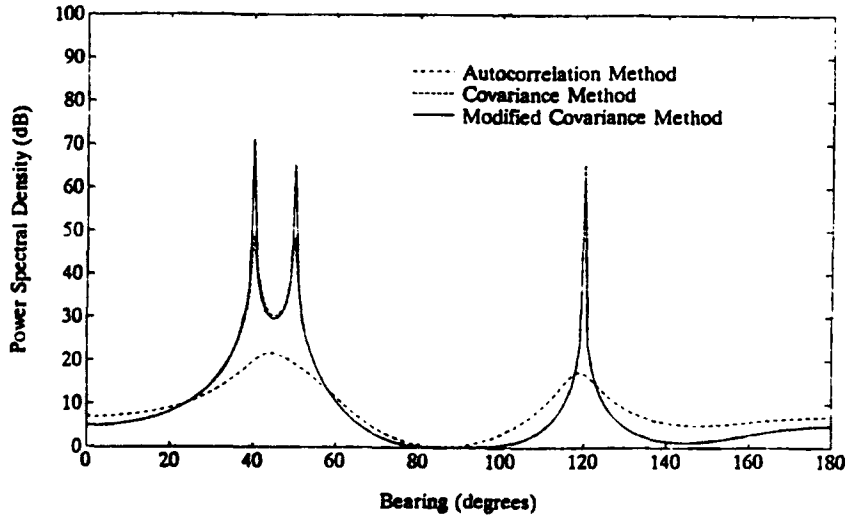


FIGURE 3.1: DF Spectrum based on three different methods of autocorrelation matrix estimation

3.3 THE COVARIANCE METHOD

Another method of autocorrelation matrix estimation is called the *covariance* method [3-1]. It operates only on the known data and thus avoids the windowing problem. Dividing the data into overlapping subarrays of size $p+1$, (a technique which is called *spatial smoothing* [3-2]) the forward data matrix is given by

$$\mathbf{X}_f = \frac{1}{\sqrt{N-p}} \begin{bmatrix} x_0, & x_1, & x_2, & \cdots, & x_{N-p-1} \\ x_1, & x_2, & x_3, & \cdots, & x_{N-p} \\ x_2, & x_3, & x_4, & \cdots, & x_{N-p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ x_p, & x_{p+1}, & x_{p+2}, & \cdots, & x_{N-1} \end{bmatrix}. \tag{3.12}$$

The corresponding backward matrix is given by,

$$\mathbf{X}_b^* = \frac{1}{\sqrt{N-p}} \begin{bmatrix} x_{N-1}, & x_{N-2}, & x_{N-3}, & \cdots, & x_p \\ x_{N-2}, & x_{N-3}, & x_{N-4}, & \cdots, & x_{p-1} \\ x_{N-3}, & x_{N-4}, & x_{N-5}, & \cdots, & x_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N-p-1}, & x_{N-p-2}, & x_{N-p-3}, & \cdots, & x_0 \end{bmatrix}. \tag{3.13}$$

In the covariance method, assuming the data is noisy and only a single sensor sample is available, the rank of the estimated autocorrelation matrix will be the smaller value of either $N-p$, which represents the number of columns of the data matrix $\mathbf{X}_f$ or $\mathbf{X}_b$, or $p+1$, which represents the number of rows of the data matrix. Ideally the rank will be greater than or equal to M for optimum performance of the superresolution estimators. Conversely, the maximum number of signal bearings that may be estimated is p subject to the constraint that this value is less than or equal to the rank of the data matrix. This constraint can be expressed as,

$$p \leq N-p \quad \text{and} \quad p \leq p+1. \tag{3.14}$$

Since the rightmost expression is always true, the largest value for p is found by solving the leftmost expression. The corresponding maximum number of signals (where $M \leq p$) then is given by,

$$M \leq \frac{N}{2}. \tag{3.15}$$

It should be noted that spatial smoothing effectively decreases the sensor aperture from N to $p+1$. This results in a decrease in resolution due to the smaller effective aperture, although this is partially offset by the averaging effect of the extra columns in the data matrix.

An example of the performance of the covariance method is shown in Figure 3.1. A complete description of this example is given at the end of section 3.3.

3.4 THE MODIFIED COVARIANCE METHOD

A more recently developed technique, called the *modified covariance* method [3-3],[3-4], doubles the number of data vectors used to form the data matrix **X**, by combining both forward and backward data matrices of the covariance method. The result is given by,

$$\mathbf{X}_{fb} = \frac{1}{\sqrt{2}} [\mathbf{X}_f, \mathbf{X}_b]. \tag{3.16}$$

where the subscripts *fb* are used to denote forward-backward, and the data matrices $\mathbf{X}_f$ and $\mathbf{X}_b$ are defined by equations (3.12) and (3.13) respectively. The modified covariance method is also sometimes called the *forward-backward* method.

In the modified covariance method, assuming the data is noisy and only a single sensor sample is available, the rank of the estimated autocorrelation matrix will be the smallest value of either $2\times(N\text{-}p)$, which represents the number of columns of $\mathbf{X}_{fb}$ (double that of $\mathbf{X}_f$ or $\mathbf{X}_b$), or $p+1$ which represents the number of rows of $\mathbf{X}_{fb}$. The maximum number of signal bearings that may be estimated is p subject to the constraint that the rank is equal to or greater than the number of signals. The constraints can be expressed as,

$$p \leq 2(N\text{-}p) \quad \text{and} \quad p \leq p+1. \tag{3.17}$$

Since the rightmost expression is always true, the largest value for p is found by solving the leftmost expression. The corresponding maximum number of signals (where $M \leq p$) then is given by,

$$M \leq \frac{2N}{3}. \tag{3.18}$$

In comparing this expression to the corresponding expression for the covariance method (equation (3.15)) the advantage of the modified covariance method is clear. A greater number of bearings can be estimated from a single sensor sample without sacrificing as much resolution due to the decreased aperture size.

An example of the performance of the modified covariance method compared to the previously described methods is given in Figure 3.1. A complete description of this example is given at the end of section 3.3.

3.5 TIME AVERAGING

In the case where a number of time samples are available, a better estimate of the autocorrelation matrix may be achieved simply by time averaging as shown here:

$$\hat{\mathbf{R}} = \frac{1}{T} \sum_{t=0}^{T-1} \hat{\mathbf{R}}(t), \tag{3.19}$$

where $\hat{\mathbf{R}}(t)$ is the autocorrelation estimate formed for time sample t. Equivalently, the data vector can be modified in the following manner,

$$\mathbf{X} = \frac{1}{\sqrt{T}} [\, \mathbf{X}(0), \mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(T\text{-}1) \,], \tag{3.20}$$

where $\mathbf{X}(t)$ is the data matrix (as described in the previous sections) formed from time sample t. In this form the relationship between time averaging is obvious, i.e. time averaging is averaging performed over time, and spatial smoothing is averaging performed over position.

The advantage of time averaging is that the resultant estimated autocorrelation matrix becomes less sensitive to the effects of temporal noise. Additionally, in the case of signals that are uncorrelated in time (i.e. they are independent signals transmitted from separate transmitters), averaging increases the number of linearly independent columns of the matrix $\mathbf{X}$ by order T compared to the submatrices $\mathbf{X}(0)$, $\mathbf{X}(1)$, etc. In this case the rank deficiency problem can be overcome simply by taking a sufficient number of time samples (i.e. $T \geq M$ using the covariance method and $T \geq M/2$ using the modified covariance method without spatial smoothing or $p = N\text{-}1$) to achieve the required rank of the estimated signal correlation matrix, without having to resort to spatial smoothing.

Finally, in the case of correlated signals (e.g. multipath) the signals do not decorrelate in time so that spatial smoothing technique must be used.

3.6 THE EFFECT OF NOISE ON AUTOCORRELATION MATRIX ESTIMATION

One way to observe the effects of noise on the estimation of the autocorrelation matrix is to determine the mean and variance of the matrix elements when they are estimated using noisy data. Starting with the covariance method with time averaging, but no spatial smoothing, the elements of the estimated autocorrelation matrix can be defined as,

$$\hat{r}_{ij} = \frac{1}{T} \sum_{t=0}^{T-1} x_i(t) x_j^*(t). \tag{3.21}$$

If the data is assumed to be corrupted by white Gaussian noise, each sensor data value can be represented as the sum of a signal plus a noise component, that is,

$$x_n(t) = s_n(t) + \eta_n(t). \tag{3.22}$$

Substituting this relationship back into equation (3.21) gives,

$$\hat{r}_{ij} = \frac{1}{T} \sum_{t=0}^{T-1} (s_i(t) s_j^*(t) + s_i(t) \eta_j^*(t) + s_j^*(t) \eta_i(t) + \eta_i(t) \eta_j^*(t)). \tag{3.23}$$

The mean value of the elements when $i \neq j$ is given by,

$$\mathrm{E}\{\hat{r}_{ij}\} = \frac{1}{T} \sum_{t=0}^{T-1} s_i(t) s_j^*(t), \tag{3.24}$$

and when $i = j$ (the main diagonal elements),

$$\mathrm{E}\{\hat{r}_{ii}\} = \sigma^2 + \frac{1}{T}\sum_{t=0}^{T-1} s_i(t)s_i^*(t), \tag{3.25}$$

where σ^2 represents the noise power.

The elemental variance is given by,

$$v = \mathrm{E}\{\,|\hat{r}_{ij} - \mathrm{E}\{\hat{r}_{ij}\}|^2\,\} = \frac{1}{T^2}(\sigma^2\sum_{t=0}^{T-1} s_i(t)s_i^*(t) + \sigma^2\sum_{t=0}^{T-1} s_j(t)s_j^*(t) + T\sigma^4). \tag{3.26}$$

The above result is based on the fact that the variance of a process YY^* is σ_y^4 and the variance of the process XY^* is $\sigma_x^2\sigma_y^2$ where X and Y represent uncorrelated white Gaussian processes with variances σ_x^2 and σ_y^2 respectively. To simplify the above variance expression a new parameter, κ_{ij}, is defined so that,

$$\kappa_{ij}s^2 = \frac{1}{2T}\sum_{t=0}^{T-1}(\, s_i(t)s_i^*(t) + s_j(t)s_j^*(t)\,), \tag{3.27}$$

where s^2 is the sum of the individual signal powers. (Note that for a large number of samples, and uncorrelated signals, $\kappa_{ij} \approx 1$). Using this definition then, the variance becomes,

$$v = \frac{2\kappa_{ij}s^2\sigma^2 + \sigma^4}{T}. \tag{3.28}$$

Since bearing accuracy is a function of the ratio of the input noise power (σ^2) to signal power (s^2), equation (3.28) can be normalized by dividing through by s^4 (which is the equivalent maximum "signal power" of the autocorrelation elements) and reexpressing the normalized variance in terms of the input signal to noise ratio. Calling this the normalized elemental variance for the covariance method, the result is given by

$$v_c = \frac{1}{T}(2\kappa_{ij}SNR^{-1} + SNR^{-2}), \tag{3.29}$$

where the input signal to noise ratio measured at a sensor is given by

$$SNR = \frac{s^2}{\sigma^2}. \tag{3.30}$$

Note that v_c^{-1} is a measure of the signal to noise power ratio of the elements of the autocorrelation matrix (as opposed to the signal to noise power ratio of the data).

Inspection of equation (3.29) shows that the normalized variance is an inverse function of SNR for signal to noise ratios greater than zero and an inverse function of SNR^2 for signal to noise ratios less than zero. Figure 3.2 illustrates this effect through simulation of a 5 element array with half wavelength spacing, $p = 4$, and $T = 5$. The bearing of the incoming signals was 40, 50, and 120 degrees, and they were uncorrelated. The variance was calculated from 1000 simulation runs performed for input signal to noise ratios ranging from -60 to +60 dB in 1 dB steps and averaged for all the elements of the autocorrelation

matrix.

The significance of this effect is that the bearing accuracy will degrade more rapidly for signal to noise ratios less than zero, indicating the need to remove as much noise as possible from the data *before* estimating the autocorrelation matrix.

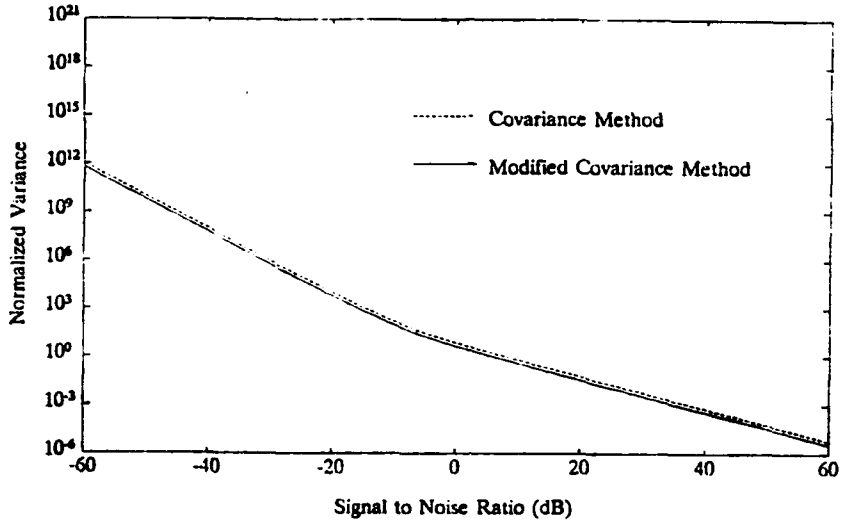


FIGURE 3.2: Elemental variance as a function of signal to noise ratio

In the more general case where spatial smoothing and/or time averaging is used equation (3.29) can easily be modified to become,

$$v_c = \frac{1}{K}(2\kappa_{ij}SNR^{-1} + SNR^{-2}). \tag{3.31}$$

Here K represents the total number of terms averaged together and can be expressed mathematically as

$$K = T(N\text{-}p). \tag{3.32}$$

If the modified covariance method is used twice as many terms are involved in the estimation of the autocorrelation matrix. For comparison purposes it is useful to keep the same expression for K and modify equation (3.31) instead. At first sight this suggests a simple relationship between the elemental variance for the covariance method (v_c) and the elemental variance for the modified covariance method (v_m), namely,

$$v_m = 0.5v_c \tag{3.33}$$

This expression is not always valid as explained in the following analysis.

A closer inspection of the modified covariance method reveals that it can be defined in terms of the covariance method as,

$$r_{ij} = 0.5(c_{ij} + c_{mn}), \tag{3.34}$$

where $m = p\text{-}j$ and $n = p\text{-}i$. In this last expression r_{ij} represents an element of the estimated autocorrelation matrix determined using the modified covariance method and c_{ij} represents an element determined using the covariance method. Since each element of the modified covariance estimate is the average of two elements of the covariance estimate, the modified covariance estimate would be expected to have half the variance (as predicted by equation (3.33)) *as long as the errors in c_{ij} and c_{mn} are uncorrelated.*

There are two conditions where the elements are correlated. The first case is for elements lying on the main cross diagonal of the autocorrelation matrix (i.e. r_{0p}, $r_{1p\text{-}1}$, $r_{2p\text{-}2}$, ..., r_{p0}). In this case c_{ij} and c_{mn} represent the same elements so that equation (3.34) simplifies to

$$r_{ij} = c_{ij}. \tag{3.35}$$

Consequently for the cross diagonal elements, the normalized variance is actually given by equation (3.31).

The second case where the elements are correlated occurs when spatial smoothing is used in conjunction with the modified covariance method. Due to the overlapping nature of the subarrays, the elements c_{ij} and c_{mn} are formed from many of the same terms (where each term has the form $x_i x_j^*$). By comparing how the two elements c_{ij} and c_{mn} are formed, an expression for the number of common terms can be derived, namely, hT where

$$h = \begin{cases} N - p - |p - i - j| & \text{if the result is} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.36}$$

Since the averaging operation in equation (3.34) has no effect on the hT common terms, then the variance will be v_c while the improvement due to the remaining $K - hT$ uncorrelated terms will be $0.5v_c$. From this, the elemental variance for the modified covariance method can be defined as,

$$v_m = \frac{v_c hT + 0.5\, v_c(K - hT)}{K} = \frac{1}{2}\left(1 + \frac{h}{N-p}\right) v_c \tag{3.37}$$

or in terms of *SNR*,

$$v_m = \frac{1}{2K}\left(1 + \frac{h}{N-p}\right)(2\kappa_{ij} SNR^{-1} + SNR^{-2}). \tag{3.38}$$

If the value of $h = 0$, then equation (3.33) applies. If on the other hand $h > 0$ then equation (3.38) can be rewritten as,

$$v_m = \frac{1}{2K}\left(2 - \frac{|p - i - j|}{N - p}\right)(2\kappa_{ij} SNR^{-1} + SNR^{-2}) \quad \text{for } h > 0. \tag{3.39}$$

Inspection of this result shows that for large values of N (i.e. $N >> 2p$), the normalized elemental variance of the modified covariance method degrades to that of the covariance method.

Given that the elemental variances are not all necessarily equal, it is useful to determine the average elemental variance since this quantity provides a more useful measure of how well the estimated autocorrelation matrix approximates the true autocorrelation matrix. To do this, the assumption is made that $\kappa_{ij} = 1$. For a large number of trials involving signals with uniformly distributed phases this assumption is reasonable. For a single trial involving a very limited number of sensor samples and/or correlated signals, the formulas derived in the following discussion will only be approximations.

In the case of the covariance method the elemental variances are all equal so that,

$$\overline{v_c} = \frac{1}{K}(2SNR^{-1} + SNR^{-2}), \tag{3.40}$$

where the overbar is used to denote the mean value. The situation is not as straight forward for the modified covariance method since the value of h in equation (3.38) changes with each element. After some algebraic manipulations, however, the result is given by,

$$\overline{v_m} = \left(\frac{4 + 3N + 3p + 5Np - N^2 - p^2}{6(p + 1)^2} \right) \overline{v_c} \qquad \text{for } p \geq \frac{N}{2} \tag{3.41}$$

and

$$\overline{v_m} = \left(1 - \frac{p(p + 2)}{6(N - p)(p + 1)}\right) \overline{v_c} \qquad \text{for } p \leq \frac{N}{2} \tag{3.42}$$

Again, as N increases for a fixed subarray size, or fixed value of p, the elemental variance of the modified covariance estimate approaches that of the covariance method.

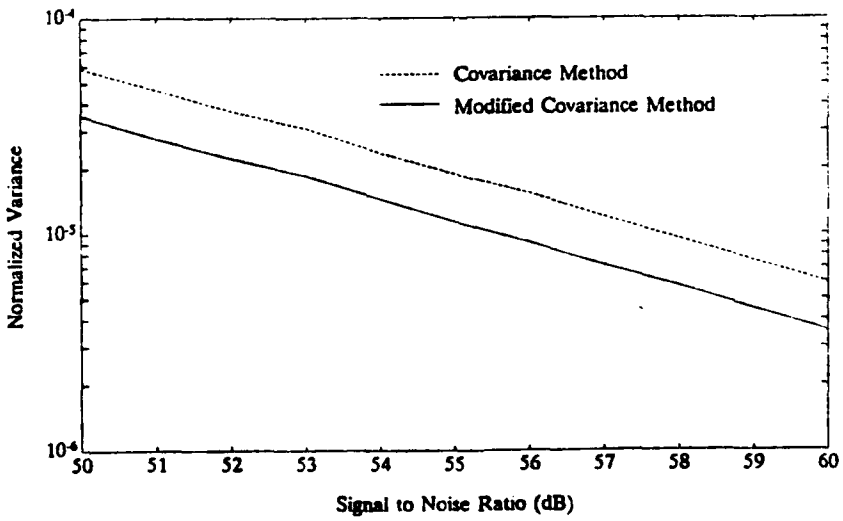


FIGURE 3.3: Blow up of the elemental variance shown in Figure 3.2

For the example shown in Figure 3.2, the predicted value of $v_m = 0.6v_c$ based on equation (3.41) which is in excellent agreement with the simulated results (see Figure 3.3). For signal to noise ratios much greater than 0, this translates into an equivalent increase in the signal to noise ratio of 2.2 dB when using the modified covariance method compared to the covariance method.

The concepts embodied by equations (3.40), (3.41), and (3.42) are illustrated in Figures 3.4 and 3.5. These figures show the decrease in the variance as a function of the number of terms for time averaging (Figure 3.4) and spatial smoothing (Figure 3.5) when either the covariance or modified covariance methods were used. In the time averaging case, the data samples were assumed to be taken from a 5 element array with half wavelength spacing. Each sensor sample was assumed to be uncorrelated with the previous sample. In the spatial smoothing case the samples were taken as overlapping subarrays (5 elements) from a single snapshot of a very large array. The bearings of the incoming signals were 40, 50, and 120 degrees and they were all uncorrelated. Statistics were computed from 1000 trials for each value of K.

As predicted by equation (3.40), the decrease in variance for the covariance method is a function of the factor $1/K$ whether time averaging or spatial smoothing was used. The same result holds true for the modified covariance method when time averaging is used as predicted by equations (3.41) and (3.42). In the spatial smoothing case the advantage of the modified covariance method over the covariance method begins to disappear as K increases, exactly as predicted by these equations. The theoretical results for Figures 3.4 and 3.5 based on the theoretical equations are not shown since they were virtually indistinguishable from the simulated results.

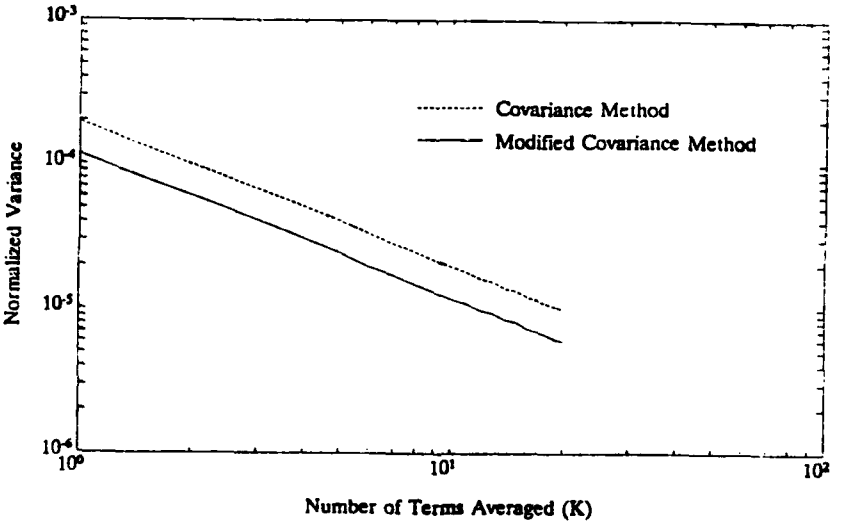


FIGURE 3.4: Normalized variance using time averaging

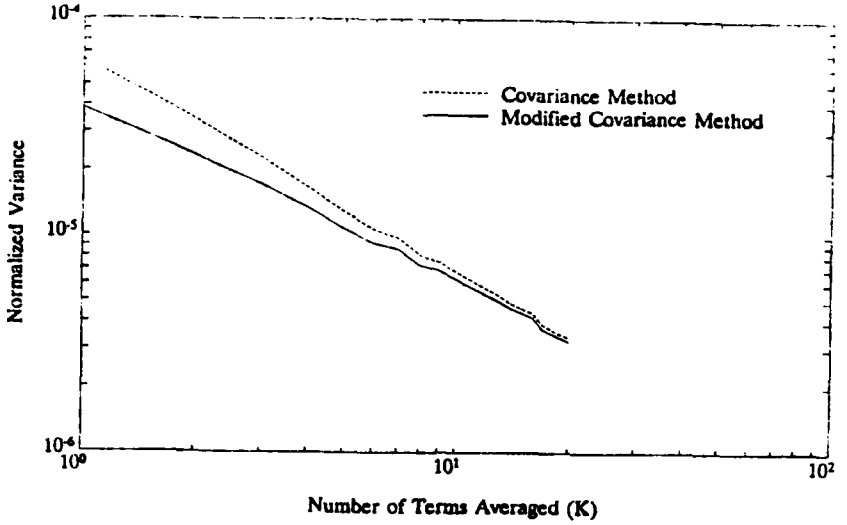


FIGURE 3.5: Normalized variance using spatial smoothing

The results presented here do not predict the ultimate bearing accuracies of any particular superresolution DF method, since the bearing estimation procedure is highly nonlinear (although linear approximations are possible at high signal to noise ratios). However, these results are useful for predicting some of the ways in which noise affects bearing estimation as well as providing a measure of the merits of various autocorrelation matrix estimation methods.

4.0 THE MODELLING APPROACH

The key to improving the performance of DF estimators over that of classical methods lies in taking better advantage of the form of the sensor data. As discussed in section 2, determining the bearing of received signals is equivalent to the problem of determining the frequencies of complex sinusoids in noise. Since the characteristics of the signal and noise are different, they can be modelled separately.

4.1 THE SIGNAL MODEL

For a single signal in a noiseless environment, the data from the n^{th} sensor in an N sensor system can be represented by equation (1.1) which is repeated here as

$$x_n = c(t)e^{-jn\omega d}. \tag{4.1}$$

An alternate representation for this equation is given by,

$$x_n = a_0 x_{n-1}, \tag{4.2}$$

where the coefficient

$$a_0 = e^{-j\omega d}, \tag{4.3}$$

and is easily computed from the data. The advantage of this alternate representation of the data is that, at least in this case, it provides a simple method of extending the data sequence and corresponding autocorrelation sequence indefinitely. The Fourier transform of the infinitely extended autocorrelation sequence then results in the ideal DF spectrum. Although the single bearing could also be determined directly from a_0, the problem becomes more difficult when several signals are involved.

In the multiple signal environment, again assuming no noise, the sensor data can be represented by,

$$x_n = \sum_{m=1}^{M} c_m(t)e^{-jn\omega_m d}, \tag{4.4}$$

where the subscript m is used to distinguish between the M signals. The equivalent alternate representation in this case is given by,

$$x_n = \sum_{m=1}^{M} a_m x_{n-m}. \tag{4.5}$$

The relationship between the coefficients represented by a_m and the spatial frequencies of the signals is not as clear cut as in the single signal case. However, a relationship does exist as demonstrated in the following analysis.

To simplify this analysis, equation (4.4) is rewritten as,

$$x_n = \sum_{m=1}^{M} c_m p_m^{-n}, \tag{4.6}$$

where for simplicity

$$c_m = c_m(t), \tag{4.7}$$

and the complex signal poles p_m are defined as

$$p_m = e^{+j\omega_m d}. \tag{4.8}$$

Since the complex amplitude c_m is of no interest for bearing determination, and noting equation (4.6) forms a linear set of equations given by,

$$\begin{aligned}
&x_0 - c_1 - c_2 - c_3 - \ldots - c_M = 0 \\
&x_1 - c_1 p_1^{-1} - c_2 p_2^{-1} - c_3 p_3^{-1} - \ldots - c_M p_M^{-1} = 0 \\
&x_2 - c_1 p_1^{-2} - c_2 p_2^{-2} - c_3 p_3^{-2} - \ldots - c_M p_M^{-2} = 0 \\
&\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\quad \vdots \\
&x_{N-1} - c_1 p_1^{-N+1} - c_2 p_2^{-N+1} - c_3 p_3^{-N+1} - \ldots - c_M p_M^{-N+1} = 0
\end{aligned} \tag{4.9}$$

then c_m can be eliminated using standard techniques. For example, to remove the first coefficient c_1 from any row (where each of the above equations is referred to as a row and are ordered as shown), the row is multiplied by p_1 and then subtracted from the previous row. If the operation is performed on the last N-1 rows, the following set of N-1 equations results:

$$\begin{aligned}
&(x_0 - x_1 p_1) - c_2(1 - p_1 p_2^{-1}) - c_3(1 - p_1 p_3^{-1}) - \ldots - c_M(1 - p_1 p_M^{-1}) = 0 \\
&(x_1 - x_2 p_1) - c_2(p_2^{-1} - p_1 p_2^{-2}) - c_3(p_3^{-1} - p_1 p_3^{-2}) - \ldots - c_M(p_M^{-1} - p_1 p_M^{-2}) = 0 \\
&(x_2 - x_3 p_1) - c_2(p_2^{-2} - p_1 p_2^{-3}) - c_3(p_3^{-2} - p_1 p_3^{-3}) - \ldots - c_M(p_M^{-2} - p_1 p_M^{-3}) = 0 \\
&\quad\vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots \\
&(x_{N-2} - x_{N-1} p_1) - c_2(p_2^{-N+2} - p_1 p_2^{-N+1}) - c_3(p_3^{-N+2} - p_1 p_3^{-N+1}) - \ldots - c_M(p_M^{-N+2} - p_1 p_M^{-N+1}) = 0
\end{aligned} \tag{4.10}$$

Removing any of the other coefficients proceeds in an identical manner. That is, to remove c_k from a row, multiply the row by p_k and subtract it from the previous row. Note that each time this operation is performed, the resultant set of equations is reduced by 1 equation.

If the procedure outlined above is carried through until all the coefficients c_m have been removed, then the resultant N-M equations have the form given by equation (4.5)

where,

$$
\begin{aligned}
a_1 &= p_1 + p_2 + p_3 + \ldots + p_M \\
a_2 &= -p_1p_2 - p_1p_3 - \ldots - p_2p_3 - \ldots - p_{M-1}p_M \\
a_3 &= p_1p_2p_3 + p_1p_2p_4 + \ldots + p_2p_3p_4 + \ldots + p_{M-2}p_{M-1}p_M \\
&\vdots \\
a_{M-1} &= (-1)^M (p_2p_3p_4 \ldots p_M + p_1p_3p_4 \ldots p_M + \ldots + p_1p_2p_4 \ldots p_{M-1}) \\
a_M &= (-1)^{M+1} (p_1p_2p_3p_4p_5 \ldots p_M)
\end{aligned}
\tag{4.11}
$$

As in the single signal case, once the coefficients represented by a_n have been determined, the data sequence x_n can be extended indefinitely. The resultant DF spectrum (computed from the extended data set using either the direct or indirect methods discussed in section 1) can be used to exactly determine the signal bearings. Figure 4.1 shows an example of the improvement in the DF spectrum using this technique compared to classical methods.

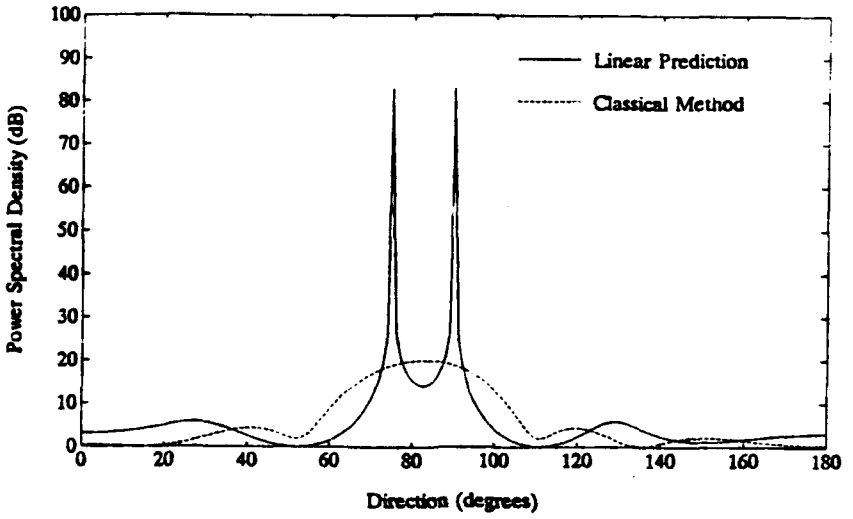


FIGURE 4.1: Comparison of the DF spectrum generated using the Linear Prediction method and the Classical method for two signals of equal power at 40 and 50 degrees.

4.2 THE NOISE MODEL

In the case where noise is present, but no signals, the sensor output may be represented by

$$x_n = \eta_n(t). \tag{4.12}$$

Since noise is not deterministic (i.e., not completely predictable), it must be handled statistically. For example, for complex white Gaussian noise, the autocorrelation sequence is given by,

$$r_{xx}(m) = 0 \quad \text{for } m \neq 0, \tag{4.13}$$

and

$$r_{xx}(0) = \sigma^2, \tag{4.14}$$

where σ^2 is the variance of the noise process. This model is useful for modelling internal sensor noise which is typically white Gaussian noise (in the temporal sense) with the same variance but uncorrelated between sensors. It is also useful for modelling external (e.g. atmospheric noise) omnidirectional noise with equal power in all directions.

For diffuse external noise sources which have an unequal noise distribution with direction, the noise can be modelled as filtered complex white Gaussian noise. That is,

$$x_n = \sum_{m=0}^{K} b_m v_{n-m}, \tag{4.15}$$

where $b_0 = 1$, K represents the order of the noise process, and v_k represents a complex Gaussian white noise process with a variance σ^2. Since v_k is not a deterministic signal, but a stochastic process, some estimation method must be used to determine the optimum value of the coefficients b_m (as opposed to determining the exact values of a_m for the signal only case discussed in the previous section).

The choice of the filter order K in equation (4.15) generally depends either on the known characteristics of the noise (e.g. for white noise $K = 0$), or is limited by the amount of available data. Since the autocorrelation sequence can only be estimated to lag N, then $K \leq N$-1.

4.3 THE SIGNAL PLUS NOISE MODEL

One approach to improving spatial frequency, which follows from the previous discussion in the preceding sections, is to combine the signal and noise models to give a model capable of handling the signals plus noise problem. One such model is called an *autoregressive moving average* (ARMA) filter and is given by,

$$x_n = -\sum_{m=1}^{p} a_m x_{n-m} + \sum_{m=0}^{q} b_m v_{n-m}, \tag{4.16}$$

Ideally the *autoregressive* (AR) filter coefficients represented by a_m are chosen to model the

signals, and the *moving average* (MA) filter coefficients represented by b_m are chosen to model the noise. In practice, this is not always true, or possible. The methods used to estimate these values are the basis of the various superresolution DF estimators discussed in this report.

4.4 COMPUTING THE DF SPECTRUM

Once the values for a_m and b_m have been estimated, the data sequence, and correspondingly the autocorrelation sequence can be extended indefinitely by computing the unknown values of x_n. The Fourier transform of the extended autocorrelation sequence then gives the power spectral density function.

Computationally, the direct method of computing the power spectral density function from the data sequence is simpler and is given by

$$S(\phi) = X(\phi)X(\phi)^*, \tag{4.17}$$

where $X(\phi)$ is the Fourier transform of the extended data sequence.

$X(\phi)$ may also be computed from equation (4.16) by taking the Fourier transform of both sides to give,

$$X(\phi) = -\sum_{m=1}^{p} a_m X(\phi)\ e^{-j\omega m} + \sum_{m=0}^{q} b_m V(\phi)\ e^{-j\omega m}. \tag{4.18}$$

and then rearranging to get

$$\mathrm{X}(\phi) = \frac{B(\phi)}{A(\phi)} V(\phi), \tag{4.19}$$

where

$$A(\phi) = 1 + \sum_{m=1}^{p} a_m\, e^{-j\omega md}, \tag{4.20}$$

and

$$B(\phi) = \sum_{m=0}^{q} b_m\, e^{-j\omega md}. \tag{4.21}$$

Substituting equations (4.3) and (4.5) back into equation (4.2), the power spectral density function can be computed based on the coefficients, a_m and b_m. That is,

$$S(\phi) = V(\phi)V^*(\phi)\frac{B(\phi)B^*(\phi)}{A(\phi)A^*(\phi)}. \tag{4.22}$$

Since for a white noise process

$$V(\phi)\,V^*(\phi) = \sigma^2, \tag{4.23}$$

where σ^2 is the variance of the noise, then

$$S(\phi) = \sigma^2 \frac{B(\phi)B^*(\phi)}{A(\phi)A^*(\phi)}. \tag{4.24}$$

The matrix representation of equation (4.24) is,

$$S(\phi) = 2\sigma^2 \frac{\mathbf{e}_q^H \mathbf{b}\mathbf{b}^H \mathbf{e}_q}{\mathbf{e}_p^H \mathbf{a}\mathbf{a}^H \mathbf{e}_p}, \tag{4.25}$$

where $\mathbf{e}_p$ and $\mathbf{e}_q$ are $p+1$ and $q+1$ element steering vectors (described in section 2.3.1 and defined by equation (2.18)), the autoregressive filter coefficient vector $\mathbf{a}$ is defined as

$$\mathbf{a} = \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \tag{4.26}$$

and the moving average coefficient vector $\mathbf{b}$ is defined as

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_q \end{bmatrix}. \tag{4.27}$$

For spectral estimation purposes, equation (4.25) is useful. However, since ultimately the goal is to determine signal bearings, a simpler form of the DF spectrum can be used. For example, the main interest in the DF spectrum is its shape (i.e. to locate the signal peaks), and consequently only the relative values of the actual spectrum are needed. Therefore the noise coefficient $2\sigma^2$ can be ignored resulting in the expression,

$$S(\phi) = \frac{\mathbf{e}_q^H \mathbf{b}\mathbf{b}^H \mathbf{e}_q}{\mathbf{e}_p^H \mathbf{a}\mathbf{a}^H \mathbf{e}_p}. \tag{4.28}$$

A further simplification can be made based on the observation that if the coefficients represented by b_m are chosen to model the noise *only*, they provide no information on the location of the signal peaks and in fact, could make it more difficult to determine the location of the true peaks by masking them. This observation does not simplify the task of calculating these coefficients, but it does result in the simplified DF spectrum given by,

$$S(\phi) = \frac{1}{\mathbf{e}_p^H \mathbf{a}\mathbf{a}^H \mathbf{e}_p}. \tag{4.29}$$

5.0 DF ESTIMATORS

In the following sections, DF estimators which are inherently based on the modelling concepts discussed in section 4, are discussed. In filter terminology these estimators may be divided into three classes, namely, all zero, all pole, and pole-zero [3-1].

5.1 ALL ZERO ESTIMATORS

In all zero estimators, or more commonly called moving average (MA) estimators, only the moving average part of the ARMA filter given by equation (4.16) model is used. The values of the autoregressive parameters a_n are set to equal 0, giving,

$$x_m = \sum_{n=0}^{q} b_n \, v_{m-n}, \tag{5.1}$$

where v_k represents a complex white Gaussian noise source. The DF spectrum can be derived from equation (4.28) by noting that the coefficients a_n all equal to zero resulting in the expression,

$$S(\phi) = \mathbf{e}_q^H \mathbf{b}\mathbf{b}^H \mathbf{e}_q. \tag{5.2}$$

From the discussion in section 4, the MA model was shown to be appropriate for modelling noise-like processes in the DF spectrum (e.g. spatial noise which has broad spectral peaks and sharp nulls). Although it was also shown in section 4 that signals can be accurately modelled using an all pole filter model, the MA model can also be used to model signals, albeit with reduced efficiency. That is, a large number of filter coefficients, compared to the numbers of signals, may be required to provide an accurate DF spectrum.

Further insight into the properties of MA estimators can be gained comparing equation (5.2) to the classical estimator defined by,

$$S(\phi) = \sum_{n=-q}^{q} \hat{r}_{xx}(n) e^{-j\omega nd}, \tag{5.3}$$

where estimation of the autocorrelation lags is discussed in section 1. Provided that the autocorrelation sequence results in a positive spectrum ($S(\phi) \geq 0$ for all ϕ) then equation (5.3) can be factored into the form given by (see also Appendix A),

$$S(\phi) = \sum_{n=0}^{q} b_n e^{-j\omega nd} \sum_{n=0}^{q} b_n^* e^{+j\omega nd}, \tag{5.4}$$

where in this case the coefficients represented by b_n are computed from the autocorrelation lags. The matrix form of this expression is identical to equation (5.2). In other words, although the underlying development philosophy is different, classical DF estimators (as described in this report) are a subclass of moving average estimators [5-1].

From this analysis, it is apparent that the performance of moving average estimators would not be expected to significantly improve on the performance of classical estimators. In general, the performance limitations of moving average estimators can be

viewed as a failure of the model to extend the data sequence beyond the known data. A consequence of this fact is that the white noise process v_k in equation (5.1) is not predictable, and so unknown values cannot be predicted.

5.2 ALL POLE ESTIMATORS

In the all pole model, only the autoregressive part of the ARMA filter defined by equation (4.16) is used. The values of the moving average coefficients b_n are set to 0, giving,

$$x_m = -\sum_{n=1}^{p} a_n x_{m-n} + v_m. \tag{5.5}$$

From the discussion in section 4, it is apparent that this model is more appropriate for generating DF spectra which contain signal peaks, than are MA techniques. As a result, the all pole estimators discussed in the following sections are generally superior, in terms of accuracy, for direction finding purposes than are MA and classical techniques. As a result, these estimators are often called superresolution DF methods. The differences in the following methods are in the manner that the filter coefficients a_n are selected (although in some of these methods a_n is not calculated directly).

5.2.1 Autoregressive Method

The Autoregressive (AR) method is based on defining the autocorrelation sequence using equation (5.5) and the relationships defined in section 2.1.1 to give,

$$r_{xx}(m) = -\sum_{n=1}^{p} a_n r_{xx}(m-n) \quad \text{for all } m>0, \tag{5.6}$$

and for $m = 0$,

$$r_{xx}(0) = -\sum_{n=1}^{p} a_n r_{xx}(-n) + \sigma^2. \tag{5.7}$$

Equations (5.6) and (5.7) are known as the Yule-Walker equations or normal equations, and are also sometimes referred to as the discrete-time Wiener-Hopf equations. Once the values of the coefficients a_n have been determined, the autocorrelation sequence can be extended infinitely, and an improved estimate of the spectrum calculated.

Equations (5.6) and (5.7) can be incorporated into a single matrix equation, called the augmented normal equations, to give,

$$\mathbf{Ra} = \sigma^2\mathbf{u}, \tag{5.8}$$

where $\mathbf{R}$ is the $(p+1)\times(p+1)$ *augmented* autocorrelation matrix, the coefficient vector $\mathbf{a}$ was defined previously by equation (4.26), and $\mathbf{u}$ is a $p+1$ element unit vector defined as,

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{5.9}$$

Assuming the autocorrelation matrix **R** is invertible (in the presence of white Gaussian noise **R** will be full rank and invertible although when estimated the result may not be), equation (5.8) can be rewritten in terms of **a** as,

$$\mathbf{a} = \sigma^2 \mathbf{R}^{-1}\mathbf{u}, \tag{5.10}$$

In cases where **R** is not invertible, the method of solution of **a** is discussed in section 7.1.

Alternatively, if only the location of spectral peaks and their power with respect to the rest of the spectrum is required, it is only necessary to solve for the coefficients a_n using the linear set of equations represented by equation (5.6). In matrix form this set of equations can be expressed by,

$$\mathbf{R}_p\mathbf{w} = -\mathbf{r}, \tag{5.11}$$

where $\mathbf{R}_p$ is the $p \times p$ *normal* autocorrelation matrix, **w** is the coefficient vector defined by

$$\mathbf{w} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix}, \tag{5.12}$$

and **r** is the vector defined by

$$\mathbf{r} = \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ \vdots \\ r_{xx}(p) \end{bmatrix}. \tag{5.13}$$

Assuming the autocorrelation matrix $\mathbf{R}_p$ is invertible, then equation (5.11) can be rewritten in terms of **w** as

$$\mathbf{w} = -\mathbf{R}_p^{-1}\mathbf{r}, \tag{5.14}$$

In cases where $\mathbf{R}_p$ is not invertible, the method of solving **w** is described in section 7.2.

The vector **r** can also be defined by noting that the augmented autocorrelation matrix **R** can be partitioned in terms of **r** and $\mathbf{R}_p$ as

$$\mathbf{R} = \begin{bmatrix} r_{zz}(0) & \vdots & \mathbf{r}^H \\ \cdots\cdots & \vdots & \cdots \\ \mathbf{r} & \vdots & \mathbf{R}_p \end{bmatrix}. \tag{5.15}$$

Although these two definitions of $\mathbf{r}$ are equivalent, the second definition is more useful when the estimated autocorrelation matrix (discussed in section 3) is used in place of the true autocorrelation matrix. In this case the true values in equation (5.15) are simply replaced by their appropriate estimates.

A simple relationship also exists between the coefficient vectors $\mathbf{a}$ and $\mathbf{w}$, namely,

$$\mathbf{a} = \begin{bmatrix} 1 \\ \cdots \\ \mathbf{w} \end{bmatrix}. \tag{5.16}$$

Once the coefficients have been determined, the power spectral density function can be calculated using equation (4.29) which is repeated here as,

$$S(\phi) = \frac{1}{\mathbf{e}_p^H \mathbf{a}\mathbf{a}^H \mathbf{e}_p}. \tag{5.17}$$

In terms of the augmented autocorrelation matrix and the solution for the vector $\mathbf{a}$ given in equation (5.9), the DF spectrum for the Autoregressive method may also be expressed as,

$$S_{AR}(\phi) = \frac{1}{\mathbf{e}_p^H \mathbf{R}^{-1}\mathbf{u}\mathbf{u}^H \mathbf{R}^{-1} \mathbf{e}_p}, \tag{5.18}$$

where the scaling factor σ^2 has been ignored.

5.2.2 Maximum Entropy Method

The Maximum Entropy (ME) method [5-2] is closely related to the Autoregressive method. In this method the extrapolation of the autocorrelation sequence is made in such a way as to maximize the entropy of the data series represented by the sequence. The data series would then be the most random, in an entropy sense, of all possible series which include the known autocorrelation lags as part of the sequence.

The entropy rate for a Gaussian random process is proportional to

$$\int_{-\omega_o}^{+\omega_o} \ln[\, S(\phi)\,]\, d\omega, \tag{5.19}$$

where the spatial frequency ω is defined in terms of the bearing ϕ by equation (1.2), $\omega_0 = \pi/d$, and the power spectral density function is represented by

$$S(\phi) = \sum_{m=-\infty}^{+\infty} r_{zz}(m)\, e^{-j\omega m d}. \tag{5.20}$$

To maximize the entropy, the derivative of equation (5.19) is taken with respect to the

unknown autocorrelation lags (i.e. $r_{xx}(m)$ where $|m| > p$ are the unknown autocorrelation lags). This leads to

$$\int_{-\omega_o}^{+\omega_o} \frac{e^{-j\omega md}}{S(\phi)}\, d\omega = 0 \quad \text{for } |m| > p. \tag{5.21}$$

Equation (5.21) implies that $\frac{1}{S(\phi)}$ has a finite Fourier expansion, that is,

$$\frac{1}{S(\phi)} = \sum_{m=-p}^{p} c_m\, e^{-j\omega md}. \tag{5.22}$$

where $c_m = c_{-m}^{*}$. The summation term on the right hand side of this expression can be factored (see Appendix A), as long as $S(\phi) \geq 0$ for all ϕ, and the resultant expression inverted to give,

$$S(\phi) = \frac{1}{\sum_{m=0}^{p} a_m e^{-j\omega md} \sum_{m=0}^{p} a_m^{*} e^{+j\omega md}}. \tag{5.23}$$

The matrix form is given by,

$$S_{ME}(\phi) = \frac{1}{\mathbf{e}_p^H \mathbf{a}\mathbf{a}^H \mathbf{e}_p}. \tag{5.24}$$

From this analysis, it is clear that the maximum entropy method belongs to the class of all pole DF estimators. Additionally it has been shown [5-3] that for the problem of signals in white Gaussian noise, and a uniform linear antenna array, the Maximum Entropy method is identical to the Autoregressive method (i.e. $S_{ME}(\phi) = S_{AR}(\phi)$). For other types of noise or antenna spacings, the two methods are not identical.

In the case of non-uniform antenna spacing, the maximum entropy solution for the set of equations represented by equation (5.21) usually requires some form of gradient search technique. This can lead to a number of practical difficulties which are not addressed in this report.

5.2.3 Linear Prediction Method

In time series modelling the Linear Prediction (LP) method predicts either a future or past data value using a sequence of current data values. In direction finding, this is equivalent to predicting either the first sensor (backward prediction), the last sensor (forward prediction), or both the first and last sensor (forward-backward prediction) in a group of sensors. These three types of predictors are described in the sections 5.2.3.1 to 5.2.3.3.

5.2.3.1 Forward Linear Prediction

Mathematically, the forward prediction case can be expressed as,

$$\hat{x}_m = -\sum_{n=1}^{p} a_{fn}\, x_{m-n}, \tag{5.25}$$

where the subscript f is used to denote the forward prediction case. The error in the estimate of x_m is given by,

$$e_{fm} = x_m - \hat{x}_m. \tag{5.26}$$

The values of the coefficients, a_{fn}, are determined by minimizing the variance of the error given by,

$$v_f = E\{|e_{fm}|^2\} = E\{e_{fm}\, e_{fm}^*\}, \tag{5.27}$$

where the error is assumed to be a zero mean process. To minimize this value, the derivative is taken with respect to each of the coefficients, a_{fk}^*, giving

$$E\{\frac{\partial v_f}{\partial a_{fk}^*}\} = 0, \tag{5.28}$$

which, using the results from equations (5.25)-(5.27) and given that $\frac{\partial a_{fk}}{\partial a_{fk}^*} = 0$, simplifies to,

$$E\{e_{fm}\, x_{m-k}^*\} = 0. \tag{5.29}$$

Replacing e_{fm} by equation (5.26), $\hat{x}_m$ by equation (5.25), and expanding gives,

$$E\{x_m\, x_{m-k}^*\} + E\{\sum_{n=1}^{p} a_{fn}\, x_{m-n}\, x_{m-k}^*\} = 0, \tag{5.30}$$

where $0 < k \leq p$. This results in a system of linear equations that can be expressed in terms of the autocorrelation parameters as,

$$r_{xx}(k) + \sum_{n=1}^{p} a_{fn}\, r_{xx}(k-n) = 0. \tag{5.31}$$

Additionally, by incorporating equation (5.26) into the right side of equation (5.27) and expanding, the optimum variance is given by

$$v_{f_{opt}} = E\{e_{fm}\, x_m^*\} - E\{e_{fm}\, \hat{x}_m^*\}. \tag{5.32}$$

Replacing $\hat{x}_m^*$ by the conjugate of equation (5.25) and then applying the result from equation (5.29) in order to simplify, then

$$v_{f_{opt}} = E\{e_{fm}\, x_m^*\}. \tag{5.33}$$

Again expanding in terms of equation (5.26),

$$v_{f_{opt}} = E\{x_m\, x_m^*\} + E\{\sum_{n=1}^{p} a_{fn}\, x_{m-n}\, x_m^*\}, \tag{5.34}$$

and re-expressing the result in terms of the autocorrelation parameters,

$$v_{f_{opt}} = r_{xx}(0) + \sum_{n=1}^{p} a_{fn}\, r_{xx}(-n). \tag{5.35}$$

The system of equations described by equations (5.31) and (5.35) can be represented in matrix form as,

$$\mathbf{R}\mathbf{a}_f = v_{f_{opt}}\mathbf{u}, \tag{5.36}$$

which is identical in form to equation (5.8), the Autoregressive model system equations. In fact, for the problem of signals in white Gaussian noise, the Linear Prediction method and the Autoregressive method are identical [5-4]. The corresponding DF spectrum then is given by equation (5.18) (i.e. $S_{LP}(\phi) = S_{AR}(\phi)$).

In the case where the true autocorrelation matrix is unknown, the matrix **R** in equation (5.36) is replaced by the estimated autocorrelation defined by,

$$\hat{\mathbf{R}} = \mathbf{X}_f\mathbf{X}_f^{\mathsf{H}} \tag{5.37}$$

where $\mathbf{X}_f$ is the forward data matrix given by one of equations (3.7), (3.10), or (3.12) (the concept of forward and backward data originated in linear prediction research [5-2],[5-5],[5-6]).

5.2.3.2 Backward Linear Prediction

The case of the backward prediction coefficients proceeds in much the same manner, where the backward estimate is defined as

$$\hat{x}_m = -\sum_{n=1}^{p} a_{bn}\, x_{m+n}, \tag{5.38}$$

and the subscript b is used to denote the backward prediction case. The final result is given by,

$$r_{xx}(-k) + \sum_{n=1}^{p} a_{bn}\, r_{xx}(n-k) = 0 \qquad \text{for } 0<k\leq p, \tag{5.39}$$

and

$$r_{xx}(0) + \sum_{n=1}^{p} a_{bn}\, r_{xx}(n) = v_{b_{opt}} \qquad \text{for } k=0. \tag{5.40}$$

The equivalent matrix solution then is,

$$\mathbf{R}^{T}\mathbf{a}_b = v_{b_{opt}}\mathbf{u}. \tag{5.41}$$

By taking advantage of the fact that the autocorrelation matrix is Hermitian, that is,

$$\mathbf{R}^{T} = \mathbf{R}^{*}, \tag{5.42}$$

and complex conjugating both sides of equation (5.41) gives

$$\mathbf{R}\mathbf{a}_b^{*} = v_{b_{opt}}\mathbf{u}. \tag{5.43}$$

Since this is identical in form to the Autoregressive equations (5.8), the expression for DF spectrum for the backward linear predictor is the same as for the Autoregressive method given by equation (5.18).

Equation (5.43) is also identical in form to equation (5.36), so that the solution for the backward coefficients and error variance can be expressed in terms of the forward values as,

$$v_{b_{opt}} = v_{f_{opt}}, \tag{5.44}$$

and

$$a_{nb} = a_{nf}^{*}. \tag{5.45}$$

As in the forward case, when the true autocorrelation matrix is unknown, an estimate is used in its place which is defined by,

$$\hat{\mathbf{R}} = \mathbf{X}_b\mathbf{X}_b^{H}, \tag{5.46}$$

where $\mathbf{X}_b$ is the backward data matrix given by one of equations (3.8), (3.11), or (3.13).

5.2.3.3 Forward-Backward Linear Prediction

In practice, when the estimated autocorrelation matrix is used, the conjugate relationship between forward and backward coefficients expressed by equation (5.45) is not true if the coefficients are calculated separately. However, by constraining these coefficients to satisfy this relationship,

$$a_{nf} = a_{nb}^{*} = a_n, \tag{5.47}$$

and solving the forward and backward prediction equations (5.36) and (5.43)

simultaneously, a better estimate of the coefficients, for spectral estimation purposes, results. This method is called the Forward-Backward Linear Prediction (FBLP) method.

The solution can be derived by minimizing the quantity $v_f + v_b$ subject to the constraint given by equation (5.47). Following the same sort of derivation as used previously in the forward or backward cases the result in matrix form is,

$$\mathbf{Ra} = v_{opt}\mathbf{u}. \tag{5.48}$$

Again the DF spectrum is given by equation (5.18) (i.e. $S_{LP}(\phi) = S_{AR}(\phi)$).

The estimated autocorrelation matrix used in the least mean square solution is the modified covariance estimate defined by,

$$\hat{\mathbf{R}} = \mathbf{X}_f\mathbf{X}_f^H + \mathbf{X}_b\mathbf{X}_b^H = \mathbf{X}_{fb}\mathbf{X}_{fb}^H, \tag{5.49}$$

where $\mathbf{X}_{fb}$ is given by equation (3.16).

5.2.4 Minimum Variance Method

The Minimum Variance (MV) method is based on the output of a beamformer which passes all energy arriving from the look direction and adaptively minimizes, in an optimal manner, the energy arriving from all other directions [5-7]. This is equivalent to minimizing the variance of the beamformer output subject to the constraint that beamformer gain is unity in the look direction.

The mathematical derivation proceeds as follows. Consider the output of a beamformer given by,

$$y_n = \sum_{k=0}^{p} c_k\, x_{n-k}, \tag{5.50}$$

where there are N antennas in the antenna array, and each output y_n is formed from a subarray of p antennas. In matrix notation the system of equations embodied by equation (5.50) can be rewritten as,

$$\mathbf{y} = \mathbf{Xc}, \tag{5.51}$$

where the output vector $\mathbf{y}$ is defined by

$$\mathbf{y} = \frac{1}{\sqrt{N-p}} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-p} \end{bmatrix}, \tag{5.52}$$

the data matrix $\mathbf{X}$ is identical to the forward data matrix $\mathbf{X}_f$ defined by equation (3.12) and the beamformer weights vector $\mathbf{c}$ is given by,

$$\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}. \tag{5.53}$$

The beamformer output variance is given by

$$v = E\{\mathbf{y}^H\mathbf{y}\}, \tag{5.54}$$

which can be rewritten in terms of the input data as,

$$v = E\{\mathbf{c}^H\mathbf{X}\mathbf{X}^H\mathbf{c}\} = \mathbf{c}^H\mathbf{R}\mathbf{c}, \tag{5.55}$$

where $\mathbf{R}$ is the autocorrelation matrix.

Assuming the antenna array is uniform and linear, the transfer function of the beamformer is given by,

$$H(\phi) = \sum_{k=0}^{p} c_k\, e^{-j\omega nd}. \tag{5.56}$$

Given that the direction of interest is represented by ϕ_0, then the beamformer gain will be constrained so that

$$H(\phi_0) = 1. \tag{5.57}$$

Expressed in matrix form,

$$\mathbf{e}^H(\phi_0)\mathbf{c} = \mathbf{c}^H\mathbf{e}(\phi_0) = 1, \tag{5.58}$$

where $\mathbf{e}(\phi_0)$ is the $p+1$ element steering vector $\mathbf{e}$ (defined by equation 2.19) evaluated at $\phi = \phi_0$.

The beamformer coefficients, c_k, are then derived by minimizing the variance defined in equation (5.55) subject to the constraint represented by equation (5.58). The solution technique for this problem is to use the Lagrange multiplier [5-8] which involves minimizing the expression

$$F = \mathbf{c}^H\mathbf{R}\mathbf{c} + \delta(\mathbf{c}^H\mathbf{e}(\phi_0) - 1), \tag{5.59}$$

with respect to the coefficient c_k^* (minimizing with respect to δ yields the original constraint equation given by equation (5.58)). Performing this minimization yields the system of equations expressed in matrix form as,

$$\mathbf{R}\mathbf{c} + \delta\mathbf{e}(\phi_0) = \underline{0}, \tag{5.60}$$

where $\underline{0}$ is a $p+1$ element column vector whose elements are all 0.

The final solution for the beamformer coefficients comes by first solving equations (5.58) and (5.60) to eliminate $\mathbf{c}$ and then expressing δ in terms of $\mathbf{e}(\phi_0)$ and $\mathbf{R}$, which gives,

$$\delta = \frac{-1}{\mathbf{e}^H(\phi_0)\mathbf{R}^{-1}\mathbf{e}(\phi_0)}, \tag{5.61}$$

then substituting this expression back into equation (5.60) and solving for $\mathbf{c}$ gives

$$\mathbf{c} = \frac{\mathbf{R}^{-1}\mathbf{e}(\phi_0)}{\mathbf{e}(\phi_0)^H\mathbf{R}^{-1}\mathbf{e}(\phi_0)}. \tag{5.62}$$

Finally, the minimum variance equation can be rewritten by incorporating this expression for $\mathbf{c}$ back into equation (5.55). The result is,

$$v_{opt} = \frac{1}{\mathbf{e}(\phi_0)^H\mathbf{R}^{-1}\mathbf{e}(\phi_0)}. \tag{5.63}$$

This equation represents the power output of the beamformer for the chosen look direction corresponding to the spatial frequency ϕ_0. The beamformer power output for any direction, then, is given by,

$$S_{MV}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}^{-1}\mathbf{e}}. \tag{5.64}$$

An important difference between this estimator and the other estimators discussed is that the Minimum Variance estimator determines the power (using equation 5.64) of the signal in the look direction, not the power density. The advantage is that the power of the signals can be determined by the height of the peaks in the DF spectrum. The disadvantage is poorer resolution (see section 5.4).

The all pole nature of this estimator is easily shown by noting that the denominator term in equation (5.64) can be factored as (see Appendix A),

$$\mathbf{e}^H\mathbf{R}^{-1}\mathbf{e} = \mathbf{e}^H\mathbf{a}\mathbf{a}^H\mathbf{e}. \tag{5.65}$$

In the case where an estimate of $\mathbf{R}$ is used, a necessary condition is $\mathbf{e}^H\hat{\mathbf{R}}^{-1}\mathbf{e} > 0$ for all ϕ (see Appendix A). The resultant estimator then has the all pole form represented by equation (4.29).

5.2.5 Thermal Noise Method

The Thermal Noise method (TN) is based on a beamformer which functions in a similar way to the Minimum Variance beamformer except the gain is not constrained to unity in the look direction [5-9]. For example, based on the previous derivation of the Minimum Variance method, the weight coefficient vector can be defined in terms of equation (5.60) to get,

$$\mathbf{c} = -\delta\mathbf{R}^{-1}\mathbf{e}, \tag{5.66}$$

where δ was originally defined by equation (5.61). If the gain constraint for the look direction is chosen to be some arbitrary nonzero function with respect to the look angle ϕ_0, equation (5.66) remains unchanged, but δ will be modified according to the new gain constraint. In other words, the purpose of δ is to control the gain of the array in the look direction.

In the Thermal Noise method the parameter δ is set to -1 with the resulting weight coefficient vector given by,

$$\mathbf{c} = \mathbf{R}^{-1}\mathbf{e}. \tag{5.67}$$

For this choice of weight vector the beamformer adaptively minimizes all energy arriving from all directions. The output in this case is noise only, that is, in equation (5.50) the beamformer output y_n represents a noise process which has been called "thermal noise".

Since this approach is very similar in concept to the Autoregressive Method, the DF spectrum is computed in an almost identical manner. That is,

$$S(\phi) = \frac{1}{H(\phi)^* H(\phi)}, \tag{5.68}$$

where $H(\phi)$ in this case is the overall beamformer transfer function (compared to equation (5.56) which is the transfer function for a particular look angle). The function $H(\phi)^*H(\phi)$ can be determined by deriving the output power spectrum in response to a white noise input with a variance of 1. For example, given the input white noise process $U(\phi)$ the output spectrum is given by,

$$V(\phi) = U(\phi)H(\phi), \tag{5.69}$$

and the output power spectrum is given by,

$$V(\phi)^* V(\phi) = U(\phi)^* U(\phi)H(\phi)^* H(\phi). \tag{5.70}$$

Since $U(\phi)$ is a white noise process with variance of 1, then $U(\phi)^* U(\phi) = 1$. Simplifying equation (5.70) then,

$$V(\phi)^* V(\phi) = H(\phi)^* H(\phi). \tag{5.71}$$

In other words for a white noise input, the output spectrum is identical to the beamformer transfer function.

The matrix form output power for equation (5.71) is given by equatior (5.55) which is repeated here as,

$$V(\phi)^* V(\phi) = \mathbf{c}^H \mathbf{R}_u \mathbf{c}. \tag{5.72}$$

Using the fact that the autocorrelation matrix for a white noise process is given by $\sigma^2\mathbf{I}$ where the variance $\sigma^2=1$, then equation (5.72) simplifies to,

$$V(\phi)^{*} V(\phi) = \mathbf{c}^{H}\mathbf{c}. \tag{5.73}$$

Finally using the results from equations (5.67), (5.68), (5.71), and (5.73), the DF spectrum for the Thermal Noise method is given by,

$$S_{TM}(\phi) = \frac{1}{\mathbf{e}^{H}\mathbf{R}^{-2}\mathbf{e}}. \tag{5.74}$$

As in the case of the MV estimator, the all pole nature of this estimator is easily shown by noting that the denominator term in equation (5.74) can be factored as,

$$\mathbf{e}^{H}\mathbf{R}^{-2}\mathbf{e} = \mathbf{e}^{H}\mathbf{a}\mathbf{a}^{H}\mathbf{e}, \tag{5.75}$$

which assumes that $S(\phi) > 0$ for all ϕ (see Appendix A). The resultant estimator then has the all pole form represented by equation (4.29).

5.3 POLE-ZERO ESTIMATORS

In section 4.3 it was shown that the autoregressive moving average (ARMA) filter was best suited for modelling signal in noise problems. Ideally the moving average (or all zero) part of the filter models the noise and the autoregressive (or all pole) part models the signal. In practice development of computationally fast algorithms based on the ideal ARMA have not been as successful as the all pole models. Current methods typically rely on using time consuming search algorithms, which are largely inappropriate for real time applications (at least until faster hardware is available). For this reason, these types of estimators are not considered in this report.

One pole-zero based method, the Adaptive Angular Response method, is considered since, as will be seen, this method is a simple modification of an all pole method.

5.3.1 Adaptive Angular Response

The Minimum Variance method can be modified to give the power density (instead of power) by dividing the Minimum Variance beamformer output by the effective noise beamwidth of the beamformer to get the average power density in the beam. In the Adaptive Angular Response (AAR) method, this technique is used to give an estimate of the true spectral power density function [5-10],[5-11]. Expressed mathematically,

$$\hat{S}(\phi) = \frac{S_{MV}(\phi)}{B_{N}(\phi)}, \tag{5.76}$$

where $S_{MV}(\phi)$ was defined previously by equation (5.64), and $B_{N}(\phi)$ is the effective noise bandwidth of the beamformer.

Evaluated for a particular look direction (i.e. $\phi = \phi_0$) the effective beamwidth can be calculated using,

$$B_N(\phi_0) = \frac{\int_0^{\pi} |H(\phi)|^2 \, d\phi}{|H(\phi_0)|^2}. \tag{5.77}$$

where the bearing angle ϕ is expressed in radians. The transfer function, $H(\phi)$, for a Minimum Variance beamformer was previously defined by equation (5.56), and can be expressed in matrix form as,

$$H(\phi) = \mathbf{c}^H \mathbf{e}. \tag{5.78}$$

Incorporating this expression into equation (5.77) and recalling that $H(\phi_0) = 1$, then

$$B_N(\phi_0) = \int_0^{\pi} \mathbf{e}^H \mathbf{c}_0 \mathbf{c}_0^H \mathbf{e} \, d\phi = \mathbf{c}_0^H \mathbf{c}_0, \tag{5.79}$$

where $\mathbf{c}_0$ is the coefficient vector $\mathbf{c}$ evaluated at $\phi = \phi_0$. Generalizing this expression for any value of ϕ and then substituting back into equation (5.76) gives,

$$\hat{S}(\phi) = \frac{S_{MV}(\phi)}{\mathbf{c}^H \mathbf{c}}, \tag{5.80}$$

Using the expression for $S_{MV}(\phi)$ given by equation (5.64) and the generalized form (i.e. any look direction) of equation (5.62) for $\mathbf{c}$, the final result is,

$$\hat{S}_{AAR}(\phi) = \frac{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}^{-2} \mathbf{e}}. \tag{5.81}$$

5.4 A COMPARISON OF DF ESTIMATORS

In this section, five different DF methods are compared which are representative of the approaches discussed up to this point. The DF estimators associated with these methods are summarized here as:

1. Bartlett (section 2.3.2): $$S_{BART}(\phi) = \mathbf{e}^H \mathbf{R} \mathbf{e} \tag{5.82}$$

2. Autoregressive (section 5.2.1):
   Maximum Entropy (section 5.2.2):
   Linear Prediction (section 5.2.3):
   $$S_{LP}(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{u} \mathbf{u}^H \mathbf{R}^{-1} \mathbf{e}} \tag{5.83}$$

3. Minimum Variance (section 5.2.4): $$S_{MV}(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}} \tag{5.84}$$

4. Thermal Noise (section 5.2.5): $$S_{TN}(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}^{-2} \mathbf{e}} \quad (5.85)$$

5. Adaptive Angular Response (section 5.3.2): $$S_{AAR}(\phi) = \frac{\mathbf{e}^H \mathbf{R}^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}^{-2} \mathbf{e}} \quad (5.86)$$

Note that for $p < N$-1, the Bartlett method becomes the Welch method [5-12].

In the simulation examples that follow (unless otherwise indicated), the signal environment consisted of three signals of equal powers and bearings of 40, 50, and 120 degrees. The direction finding array consisted of 8 colinear sensors with half wavelength spacing. Noise between sensors was uncorrelated. Signal phases and sensor noise were also uncorrelated from trial to trial where each trial consisted of estimating the bearings as the signal to noise ratios were varied from 5 to 65 dB in 1 dB steps (the noise was scaled accordingly for each step). Bearing error variance statistics were based on 300 such trials.

Bearing errors were calculated by determining the bearing of the three largest peaks in the spectrum and subtracting these values from the corresponding true values. Bearing accuracy was then calculated as the root mean squared value of these errors. Justification for choosing the three largest peaks in the spectrum is based on the fact that the heights of the peaks are generally related to the square of the corresponding signal power (except for the Bartlett and Minimum Variance methods where the peaks are proportional) [5-4],[5-9]. Although this is not necessarily the optimal method for choosing the correct signal peaks, it is useful for comparing various methods and highlighting the problems involved.

Since multipath is considered to be the most significant problem in tactical DF at VHF/UHF, the autocorrelation matrix estimates were generated from single sample samples (i.e. $T = 1$) using spatial smoothing. The choice of the model order p for spatial smoothing purposes was mainly based on choosing the model order for which estimator accuracy was the highest for a given signal to noise ratio as shown in Figure 5.1. The variance of the bearing estimates for each value of p were computed from 300 trials at a fixed signal to noise ratio of 50 dB.

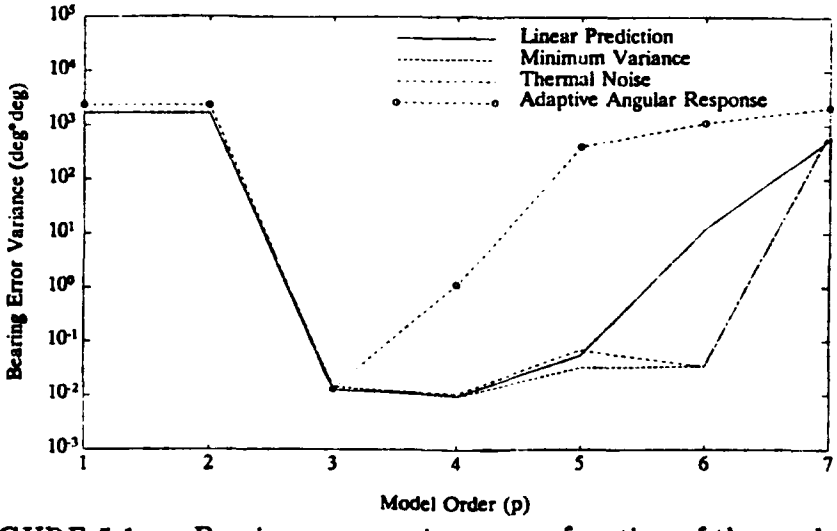


FIGURE 5.1: Bearing error variance as a function of the model order p

For the Linear Prediction, and Minimum Variance methods the optimum choice based on Figure 5.1 was $p = 4$, and for the Adaptive Angular Response method the optimum choice was $p = 3$. Although the optimum choice for the Thermal Noise method using Figure 5.1 is $p = 4$, it was found in simulations that this method performed better at lower signal to noise ratios for a model order of $p = 3$ with only a marginal decrease in accuracy. For this reason, a model order of $p = 3$ was used in simulations involving the Thermal Noise method. (Note model order selection for situations where the true signal bearings are unknown is discussed in section 10).

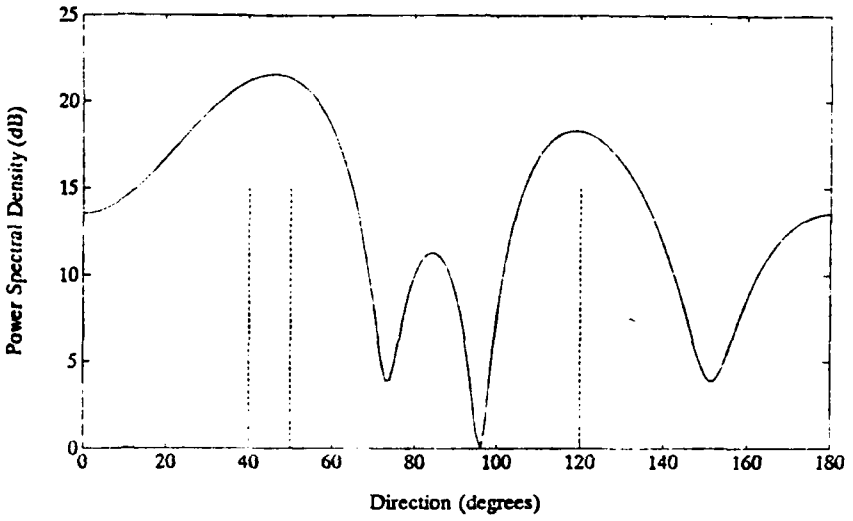


FIGURE 5.2: DF Spectrum of the Bartlett method for the noiseless case (dashed lines show the true signal bearings).

Figure 5.2 provides an example of the DF spectrum generated using the Bartlett estimator in a noiseless environment. As mentioned before, the resolution of classical and moving average methods is poor as illustrated in this example (i.e. the bearings at 40 and 50 degrees are unresolved). For this reason, the Bartlett estimator is not considered in the following simulation examples.

Of the matrix estimation schemes, the results in section 3.6 implied that under the same conditions, the modified covariance method was superior to the covariance method. These results are confirmed in Figure 5.3 which illustrates the improvement in the Thermal Noise estimator accuracy versus signal to noise ratio using the two different methods. In this particular example, the c' variance method produced results which were about 10 dB poorer in terms of the signal to noise ratio than the modified covariance method. This is worse than would be predicted from the results given in section 3.6 and probably due to the fact that the autocorrelation matrix estimate for the covariance method was not full rank which degrades the results somewhat. Results using other estimators are similar. For these reasons, all further comparisons between estimators in this and later sections is done using the modified covariance method.

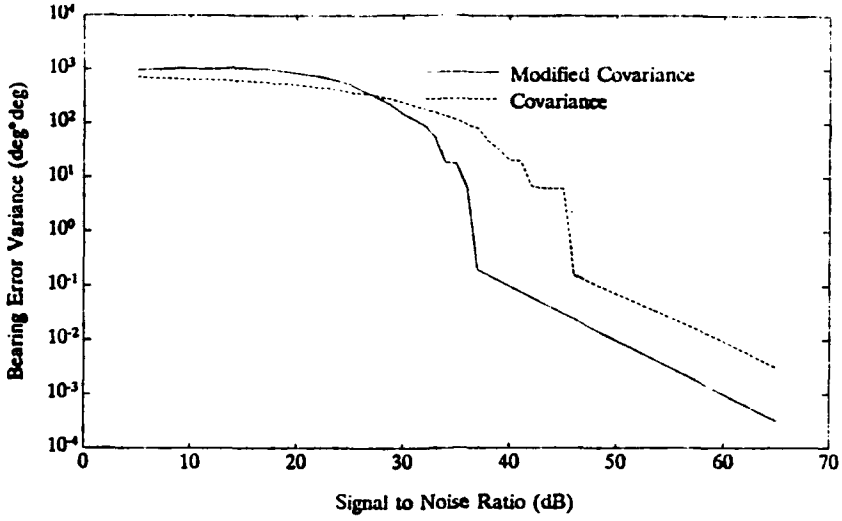


FIGURE 5.3: Comparison of the Covariance and Modified Covariance methods used in conjunction with the Minimum Variance method.

Figure 5.4 compares the performance of 4 DF estimators. From the results shown it is clear that at high signal to noise ratios the estimators are well behaved with the bearing error variance inversely proportional to the signal to noise ratio. In terms of the mean elemental variance of the estimated autocorrelation matrix (described in section 3.6), the bearing variance is directly proportional to the elemental variance. This is in keeping with the comment that at high signal to noise ratios estimator performance will be approximately linear.

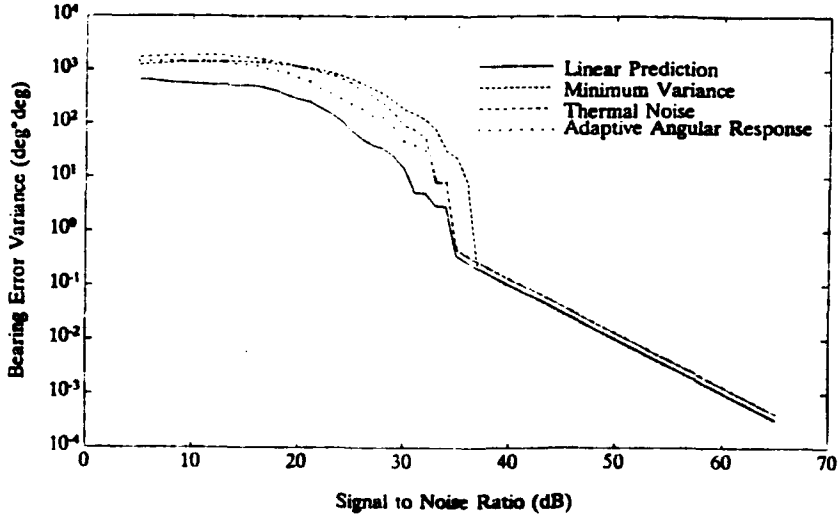


FIGURE 5.4: Comparison of DF estimators

In terms of estimator accuracy, at high signal to noise ratios, the Linear Prediction and Minimum Variance methods had the best performance. At lower signal to noise ratios, estimator performance in Figure 5.4 departs dramatically from their linear behaviour. The point at which this failure occurs is called threshold. In terms of threshold effects, the best performance was achieved equally by the Adaptive Angular Response, Linear Prediction, and Thermal Noise methods, while the Minimum Variance method had the poorest performance. In all of these methods for this simulation, the threshold effect was caused by merging of the signal peaks at 40 and 50 degrees in the DF spectrum. Consequently the lower the threshold, the better the resolution of the method.

Although the Adaptive Angular Response method performed reasonably well in these simulations, this method was extremely sensitive to model order as Figure 5.1 indicates. At higher model orders ($p > 3$) the performance of this method was extremely poor (e.g. see Figure 5.1). This is due to two effects, spurious peaks in the DF spectrum which are mistaken for true signal peaks, and spectral peak inversion. The spectral peak inversion problem is illustrated in Figure 5.5. To understand the cause, it is useful to note that the spectrum of the Adaptive Angular Response estimator is simply the ratio of the Thermal Noise DF spectrum divided by the Minimum Variance DF spectrum at any particular bearing. Normally the peaks in the Thermal Noise spectrum are larger than the corresponding peaks in the Minimum Variance spectrum so that the resultant Adaptive Angular Response spectrum has positive peaks. Occasionally the opposite is true with the result that the resultant peaks are inverted, i.e., valleys are formed where the peaks should be. Given these problems, the choice of model order for the Adaptive Angular Response method appears to be limited to $p = M$ which ensures that a maximum of only M peaks will exist in the DF spectrum (no spurious peaks), and at least for the simulations performed for Figure 5.4, avoids the spectral inversion problem.

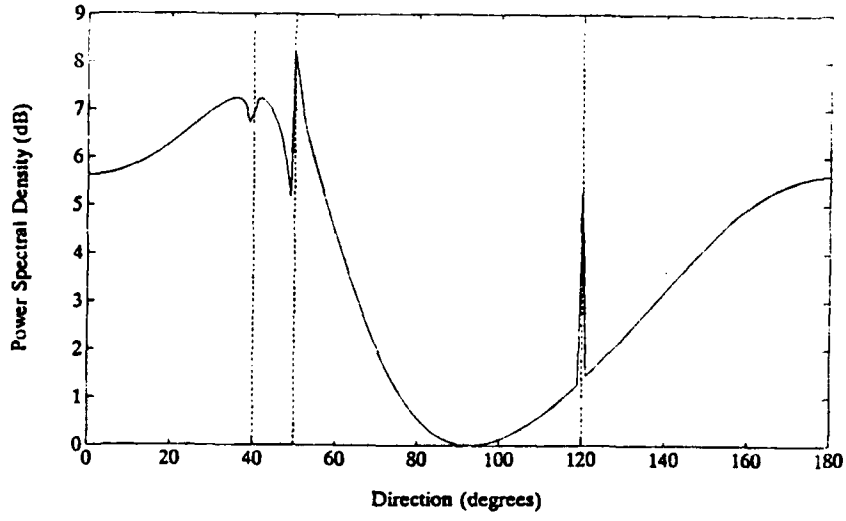


FIGURE 5.5: Spectral peak inversion in the Adaptive Angular Response DF spectrum (dashed lines show the true signal bearings)

The form of the Thermal Noise estimator (equation (5.85)) provides a useful basis of comparison with the other estimators. For example the only difference between the Thermal Noise estimator and the Minimum Variance estimator (equation 5.84) is the power to which the inverse autocorrelation matrix is raised. Squaring has the effect of making the peaks in the DF spectrum more pronounced which potentially improves the resolution capabilities of the method. This explains the poor performance of the Minimum Variance method for $p = 3$ in Figure 5.1 where in a number of trials the signals at 40 and 50 degrees were not resolved. Raising the inverse autocorrelation matrix to higher powers is possible, but this also has the effect of emphasizing spurious peaks which can degrade performance of the estimator. Figure 5.6 illustrates these effects for various powers using the estimator defined by

$$S(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}^{-m} \mathbf{e}} \tag{5.87}$$

where $m = 1$ for the Minimum Variance method and $m = 2$ for the Thermal Noise method. The generated spectrums have also been offset for clarity.

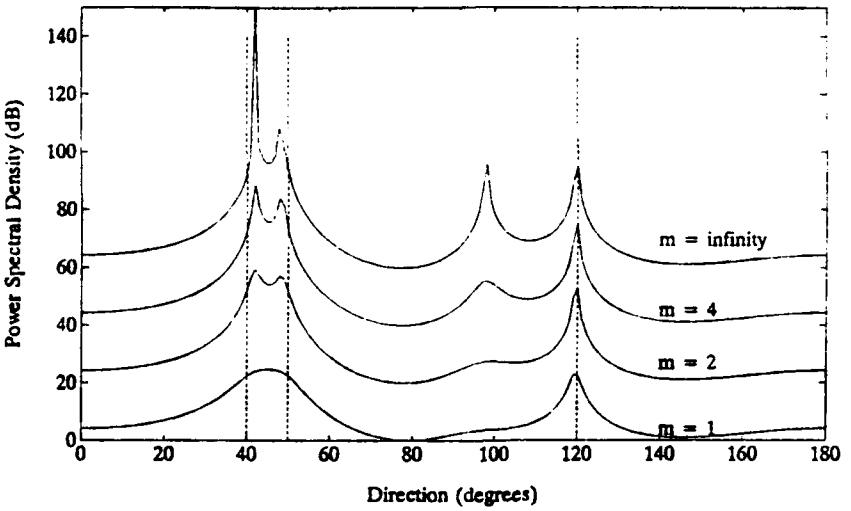


FIGURE 5.6: Spectrum of the DF estimator defined by equation (5.87) using various power of m (dashed lines show the true signal bearings).

As mentioned earlier, the choice of model order for the Thermal Noise method in Figure 5.4 was $p = 3$, even though accuracy was better for $p = 4$ based on the results shown in Figure 5.1. It was found, however, that in simulations where $p = 4$ the threshold of the Thermal Noise method degraded to that of the Minimum Variance method shown in Figure 5.4. At higher orders the accuracy decreased significantly and spurious estimates were a problem.

In comparing the Linear Prediction method to the Thermal Noise method, it is useful to rewrite equation (5.85) for the Thermal Noise estimator as,

$$S_{TN}(\phi) = \frac{1}{\mathbf{e}^H\left(\sum_{i=0}^{p} \mathbf{R}^{-1}\mathbf{u}_i\mathbf{u}_i^H\mathbf{R}^{-1}\right)\mathbf{e}}, \tag{5.88}$$

where $\mathbf{u}_i$ is a $p+1$ element column vector of zeros with a 1 in the i^{th} position. In this form the Thermal Noise method is clearly related to the Linear Prediction method (equation (5.83)) except whereas the Linear Prediction method is based on using a single prediction filter ($i = 0$ for forward prediction and $i = p$ for backward prediction) the Thermal Noise method combines the power outputs from $p+1$ filters where the i^{th} filter predicts (or interpolates) the i^{th} element in the sensor subarray being processed. The outputs from each of these filters is also inversely weighted according to the prediction error variance, i.e., the better the filter model fit with the data the greater the weighting.

In equation (5.88) each of the p filters can be determined independently. Equation (5.85) then performs an average of the weighted filter outputs. This is different from the idea of forward-backward linear prediction which constrains the forward and backward filter coefficients to be related before the coefficients are calculated.

For comparison purposes it is also useful to define the null spectrum as,

$$D(\phi) = \frac{1}{S(\phi)}, \tag{5.89}$$

so-called since nulls in $D(\phi)$ correspond to peaks in the DF spectrum. The null spectrum of the Thermal Noise method can also be interpreted as the average of the null spectrums of the $p+1$ individual prediction/interpolation filters. A disadvantage of the Thermal Noise method is that in cases where the estimated autocorrelation matrix is used, each prediction/interpolation filter will produce nulls in its own null spectrum at slightly different bearings. Averaging the nulls together, unless they are exactly aligned, decreases the overall null depth resulting in smaller peaks in the DF spectrum and poorer resolution compared to the case where only a single filter is used (i.e. linear prediction). An advantage, however, is that averaging also increases the immunity to spurious peaks since the corresponding spurious nulls in the individual null spectrums often don't occur at the same bearing and as a result the corresponding spurious peaks are smoothed out in the DF spectrum.

Ideally averaging the null spectrums in the Thermal Noise method should also lead to more stable estimates. In practice it has been found that the error variance and spurious performance of the DF estimator defined by,

$$S(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}^{-1}\mathbf{u}_i\mathbf{u}_i^H\mathbf{R}^{-1}\mathbf{e}}, \tag{5.90}$$

is a nonlinear function of $|i - 0.5p|$. That is, the accuracy and suppression of spurious peaks is best for the linear prediction case where the outside data values in the sensor subarray are being predicted ($i = 0$ or $i = p$) and is worst when the middle data value(s) of the sensor subarray are being interpolated ($i = 0.5p$ or $i = 0.5(p+1)$). As a result, the accuracy of the Thermal Noise method was slightly poorer than the Linear Prediction method.

An example of these effects is shown in Figure 5.7 where the DF spectrum of the Thermal Noise method is compared to the spectrums generated using equation 5.90. Each of the DF spectrums have been offset for clarity, and the spectrums corresponding to $i = 3$, and $i = 4$ have not been plotted since these are identical to the spectrums corresponding to $i = 0$ and $i = 1$ respectively. In comparing the various spectrums, the spectrum for $i = 2$ has the poorest accuracy (i.e. the peaks corresponding to the 40 and 50 degree signal bearings exhibit greater error) and also contains a large spurious peak at 110 degrees (although this has little effect on the Thermal Noise spectrum). In comparison, the spectrum for $i = 0$, the Linear Prediction spectrum, has the best accuracy.

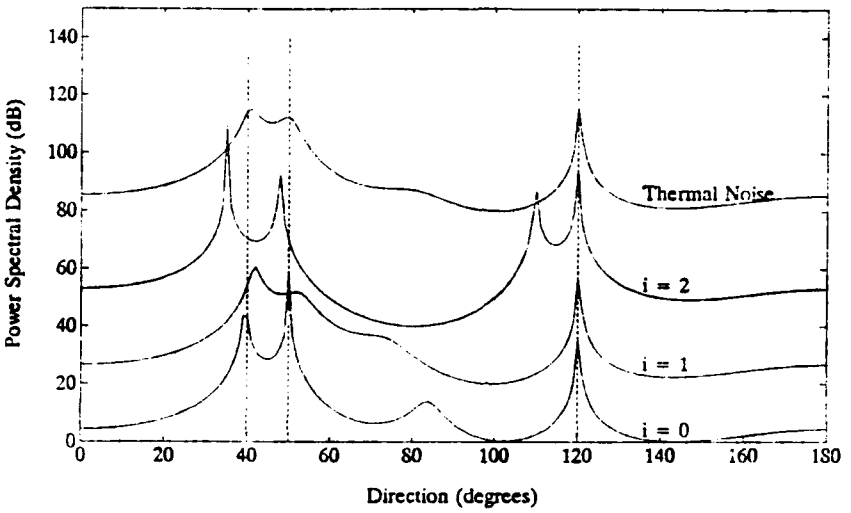


FIGURE 5.7: DF spectrums for the Thermal Noise method and the corresponding spectrums of the prediction/interpolation filters (dashed lines show the true signal bearings).

Based on the results of comparisons between the various DF estimators, two criteria are clearly important in determining estimator performance. The first is estimator accuracy above threshold, and the second is the signal to noise ratio where threshold occurs. In terms of accuracy, methods which operate using higher model orders achieve better accuracy above threshold. The signal to noise ratio at which threshold occurs is a function of estimator resolution - the better the resolving abilities of the estimator the lower the threshold (where two signals close in bearing are considered resolved when two corresponding closely spaced peaks are formed in the DF spectrum). In terms of this criteria, the Linear Prediction method had the best performance for the simulations summarized by Figure 5.4.

6.0 SIGNAL AND NOISE SUBSPACE DIVISION

In many practical situations, the sensor data is the measurement of a signal or signals corrupted by noise. Since the signal spectrum is of primary interest, the addition of noise only clouds the issue. The all pole type spectral estimators discussed to this point (which include AR, LP, ME, MV, and TN) are based on models which do not specifically address this issue, so consequently, as the input signal to noise ratio decreases, the spectral estimator performance degrades.

An example of this can be illustrated by considering the problem of direction finding on several point sources when the sensor data is corrupted by uncorrelated complex white Gaussian noise. The resultant sensor data can be represented by,

$$x_m = s_m + n_m. \tag{6.1}$$

Here s_m represents the uncorrupted sensor inputs (i.e. the sensor inputs in the absence of noise) which can be modelled exactly by an all pole model (see Section 4.1) given as,

$$s_m = -\sum_{n=1}^{p} c_n \, s_{m-n}. \tag{6.2}$$

The process n_m represents uncorrelated complex white Gaussian noise with variance η^2. The exact spectrum of x_m is given by,

$$S(\phi) = \frac{1}{C(\phi)C^*(\phi)} + \eta^2, \tag{6.3}$$

or alternatively,

$$S(\phi) = \frac{1 + \eta^2 C(\phi)\, C^*(\phi)}{C(\phi)C^*(\phi)}, \tag{6.4}$$

where,

$$C(\phi) = 1 + \sum_{n=1}^{p} c_n \, e^{-j\omega nd}. \tag{6.5}$$

Clearly, from the form of equation (6.4), $S(\phi)$ is the spectrum of an ARMA process. That is, the addition of complex white Gaussian noise to an all pole process results in an ARMA process.

From the preceding analysis it is apparent that the ARMA spectral estimator would be the most appropriate choice for estimating the spectrum of x_m. However, as mentioned before, this results in a difficult non-linear search problem, so all pole estimators are often used instead. Using any of the all pole methods, the estimated spectrum has the form,

$$\hat{S}(\phi) = \frac{1}{A(\phi)A^*(\phi)}. \tag{6.6}$$

In comparing the expression for the estimated spectrum to the expression for the true spectrum, equation (6.4), it is quite clear that the estimated spectrum will only approach the true spectrum as the additive noise power, η^2, approaches 0. Conversely, the addition of noise degrades the estimated spectrum.

If a method could be developed which would remove most of the noise before processing with an all pole spectral estimator, the estimated DF spectrum would be greatly improved. A relatively simple approach is based on the observation that the autocorrelation sequence of the sensor data x_m described by equation (6.1), is given by

$$r_{xx}(m) = E\{x_{n+m}\, x_n^*\}. \tag{6.7}$$

Expanding in terms of s_m and n_m,

$$r_{xx}(m) = E\{(s_{n+m} + n_{n+m})(s_n + n_n)^*\}. \tag{6.8}$$

Since s_m and n_m are uncorrelated then,

$$r_{xx}(m) = E\{s_{n+m}\, s_n^*\} + E\{n_{n+m}\, n_n^*\}. \tag{6.9}$$

Based on this last equation, the autocorrelation matrix can be expressed as,

$$\mathbf{R} = \mathbf{R}_s + \mathbf{R}_n, \tag{6.10}$$

where $\mathbf{R}_s$ is the signal autocorrelation matrix and $\mathbf{R}_n$ is the noise autocorrelation matrix. In the special case where n_m represents an additive white noise process with a variance η^2, then the autocorrelation sequence can be expressed as,

$$r_{xx}(0) = E\{s_n\, s_n^*\} + \eta^2, \tag{6.11}$$

which represents the signal power plus the noise power, and

$$r_{xx}(m) = E\{s_{n+m}\, s_n^*\} \qquad \text{for } m \neq 0. \tag{6.12}$$

The matrix equation (6.10) becomes,

$$\mathbf{R} = \mathbf{R}_s + \eta^2\mathbf{I}, \tag{6.13}$$

where $\mathbf{I}$ is the identity matrix.

From these results, the effects of the noise can be removed and the DF spectrum estimate enhanced by subtracting the value of η^2 from either $r_{xx}(0)$ in the autocorrelation sequence, or from each element in the principle diagonal of the autocorrelation matrix. The difficulty with this method is in determining the value of the noise power, η^2. Additionally, in practice due to limited data, only an estimate of the autocorrelation matrix is available in which case equations (6.11) to (6.13) will only

approximately be true. In this case noise will affect not only $r_{xx}(0)$, but all other estimated autocorrelation lags as well.

6.1 EIGENANALYSIS

Further insight into the noise problem can be gained by examining the structure of the autocorrelation matrix in more detail. For the case of M signals, the signal autocorrelation matrix can be expressed as,

$$\mathbf{R}_s = \sum_{m=0}^{M} \rho_m \mathbf{s}_m \mathbf{s}_m^H \tag{6.14}$$

where ρ_m represents the signal power of the m^{th} incident signal, and the normalized signal vectors each have the form given by

$$\mathbf{s}_m = \frac{1}{\sqrt{N}} \left[1, \; e^{+j\omega_m d}, \; e^{+j\omega_m 2d}, \; \ldots, \; e^{+j\omega_m (N-1)d} \right]^T. \tag{6.15}$$

The parameter ω_m is the spatial frequency corresponding to the bearing of the m^{th} signal. Since each matrix formed from the product of the individual signal vectors, $\mathbf{s}_m \mathbf{s}_m^H$, has rank 1, the signal correlation matrix, which is the sum of these matrices, will have rank equal to the number of signals, M, or have the full rank $p+1$ if $M > p$. Since the signal bearings can only be uniquely solved if $M \leq p$ (i.e. at least M equations formed from the rows or columns of $\mathbf{R}$ are are required to uniquely solve for the M unknown bearings), then only this case is considered here.

One method of examining the noise effects is based on decomposing the signal autocorrelation matrix into its component eigenvalues and eigenvectors. The eigenvalues, λ_i, and corresponding eigenvectors, $\mathbf{v}_i$, of a square matrix, $\mathbf{M}$, have the property that,

$$\mathbf{M}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{6.16}$$

where the quantity $\mathbf{v}_i^H \mathbf{v}_i = 1$. If the matrix is Hermitian (i.e. $\mathbf{M}^H = \mathbf{M}$), then the eigenvalues will be real, and if in addition, the eigenvalues are all distinct, the eigenvectors will form an orthonormal set. The methods used to compute the eigenvalues and eigenvectors are not discussed in this report but can be found in reference [6-1].

In the case of the signal autocorrelation matrix the eigenvector expression becomes,

$$\mathbf{R}_s \mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{6.17}$$

The eigenvectors form an orthonormal basis set for the signal vectors, so that the matrix $\mathbf{R}_s$ can be decomposed in terms of all its eigenvalues and eigenvectors as,

$$\mathbf{R}_s = \sum_{m=0}^{p} \lambda_m \mathbf{v}_m \mathbf{v}_m^H, \tag{6.18}$$

in which the eigenvalues have been ordered in decreasing value ($\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_p$), and the eigenvectors associated with the non-zero eigenvalues are orthonormal (i.e., $\mathbf{v}_i^H \mathbf{v}_j = 1$ if

$i = j$, and $\mathbf{v}_i^H \mathbf{v}_j = 0$ if $i \neq j$). The remaining eigenvectors (if more than one), which are associated with the zero eigenvalues can arbitrarily chosen to satisfy equation (6.16) as long as the vector chosen is orthogonal to all the eigenvectors associated with the non-zero eigenvalues (but not necessarily those associated with the zero eigenvectors). For simplicity, however, these eigenvectors are assumed (throughout the rest of this report) to be chosen to satisfy the same orthonormal condition as the eigenvectors associated with the non-zero eigenvectors.

Since the eigenvectors associated with non-zero eigenvalues span the same subspace as the signal vectors, there will only be M non-zero eigenvalues, that is, $\lambda_{M+1} = \lambda_{M+2} = \ldots = \lambda_p = 0$. Equation (6.18) can be rewritten as,

$$\mathbf{R}_s = \sum_{m=0}^{M-1} \lambda_m \mathbf{v}_m \mathbf{v}_m^H, \tag{6.19}$$

The first M eigenvectors are known as the principal eigenvectors.

Similarly the decomposition of the autocorrelation matrix for the case of signals in white noise can be represented by

$$\mathbf{R} = \sum_{m=0}^{p} \psi_m \mathbf{u}_m \mathbf{u}_m^H, \tag{6.20}$$

where the eigenvalues ψ_i and eigenvectors $\mathbf{u}_i$ satisfy the relationship

$$\mathbf{R}\mathbf{u}_i = \psi_i \mathbf{u}_i. \tag{6.21}$$

Breaking the autocorrelation matrix down into the signal and noise autocorrelation matrices (from equation (6.13)) then,

$$(\mathbf{R}_s + \eta^2 \mathbf{I})\mathbf{u}_i = \psi_i \mathbf{u}_i. \tag{6.22}$$

From this expression it is apparent that the autocorrelation matrix will be full rank (i.e. $p+1$ non-zero eigenvectors) since the noise autocorrelation matrix, $\eta^2\mathbf{I}$, has full rank.

The relationship between the eigenvalues and eigenvectors for the autocorrelation matrix, and those of the signal autocorrelation matrix can be highlighted by rearranging equation (6.22) as,

$$\mathbf{R}_s \mathbf{u}_i = (\psi_i - \eta^2)\mathbf{u}_i. \tag{6.23}$$

This is, in fact, another form of the eigenvector expression for the signal autocorrelation matrix given by equation (6.17). Therefore,

$$\mathbf{u}_i = \mathbf{v}_i, \tag{6.24}$$

and

$$\psi_i = \lambda_i + \eta^2. \tag{6.25}$$

Using these results, equation (6.20) can be expressed in terms of the signal autocorrelation eigenvalues and eigenvectors as,

$$\mathbf{R} = \sum_{m=0}^{p} \psi_m \mathbf{v}_m \mathbf{v}_m^H \tag{6.26}$$

which can be rewritten as,

$$\mathbf{R} = \sum_{m=0}^{M-1} (\lambda_m + \eta^2) \mathbf{v}_m \mathbf{v}_m^H + \sum_{m=M}^{p} \eta^2 \mathbf{v}_m \mathbf{v}_m^H. \tag{6.27}$$

This represents the decomposition of the autocorrelation matrix into its associated eigenvalues and eigenvectors. The first summation term of equation (6.27) contains the eigenvectors that span the signal subspace (i.e. each of these vectors is a linear combination of the signal vectors), and in this context the first M eigenvectors are said to form the signal subspace. The remaining p-M+1 eigenvectors are orthogonal to the signal subspace vectors and are said to form the noise subspace.

In practice, since only an estimate of the autocorrelation matrix is available, the above analysis is only approximately true. In this case, and given no knowledge about the noise, the left summation term of equation (6.27) represents the optimum reduced rank approximation in the least squares sense of the signal autocorrelation matrix [6-2] (see also Appendix B). If the noise is known to be white Gaussian in nature, an even better estimate of the signal correlation matrix can be obtained by subtracting an estimate of the η^2 noise contribution from the largest M eigenvalues where the estimate of η^2 is obtained by the averaging the p-M+1 smallest eigenvalues. Since removal of the η^2 only affects the shape of the estimated spectrum, not the location of the peaks, this additional step is not normally performed in practice.

6.2 SINGULAR VALUE DECOMPOSITION

Division of the autocorrelation matrix into signal and noise subspaces can be accomplished using eigenanalysis techniques as discussed in the last section. A similar division may also be made in the data matrix using singular value decomposition (SVD) techniques. According to the SVD theorem [6-3], an arbitrary $m \times n$ complex valued matrix $\mathbf{A}$ of rank K can be decomposed in terms of the orthonormal left singular vectors, $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{K-1}$, the orthonormal right singular vectors, $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{K-1}$, and the positive real singular values, $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{K-1}$, as,

$$\mathbf{A} = \sum_{i=0}^{K-1} \sigma_i \mathbf{u}_i \mathbf{v}_i^H. \tag{6.28}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are m and n element vectors respectively, and the singular values are arranged in decreasing order (i.e. $\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{k-1} > 0$). The relationship between the singular vectors and the matrix $\mathbf{A}$ can also be expressed as,

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \quad \text{and} \quad \mathbf{A}^H \mathbf{u}_i = \sigma_i \mathbf{v}_i. \tag{6.29}$$

The relationship between SVD and eigenanalysis can be shown through the

following derivation:

$$\begin{aligned}
\mathbf{A}\mathbf{v}_i &= \sigma_i \mathbf{u}_i \\
(\mathbf{A}\mathbf{v}_i)^H \mathbf{A}\mathbf{v}_i &= (\sigma_i \mathbf{u}_i)^H \sigma_i \mathbf{u}_i \\
\mathbf{v}_i^H \mathbf{A}^H \mathbf{A}\mathbf{v}_i &= \sigma_i^2 \mathbf{u}_i^H \mathbf{u}_i \\
\mathbf{v}_i^H \mathbf{A}^H \mathbf{A}\mathbf{v}_i &= \sigma_i^2 \mathbf{v}_i^H \mathbf{v}_i \qquad (\text{since } \mathbf{u}_i^H \mathbf{u}_i = \mathbf{v}_i^H \mathbf{v}_i = 1) \\
\mathbf{A}^H \mathbf{A}\mathbf{v}_i &= \sigma_i^2 \mathbf{v}_i . \qquad (6.30)
\end{aligned}$$

Similarly,

$$\mathbf{A}\mathbf{A}^H \mathbf{u}_i = \sigma_i^2 \mathbf{u}_i . \qquad (6.31)$$

From the above two relationships, it is apparent that the vectors $\mathbf{v}_i$ are eigenvectors of the matrix $\mathbf{A}^H\mathbf{A}$ and the vectors $\mathbf{u}_i$ are eigenvectors of the matrix $\mathbf{A}\mathbf{A}^H$. The singular values σ_i are the positive square root of the corresponding non-zero eigenvalues of either matrix.

If SVD is performed on the signal data matrix then the singular value decomposition of this matrix has the form,

$$\mathbf{X}_s = \sum_{m=0}^{M-1} \sigma_m \mathbf{u}_m \mathbf{v}_m^H . \qquad (6.32)$$

For the case of signals in white noise, a similar decomposition of the data matrix can be performed which results in,

$$\mathbf{X} = \sum_{m=0}^{p} \delta_m \mathbf{w}_m \mathbf{z}_m^H . \qquad (6.33)$$

where δ_m represents the singular values, and $\mathbf{w}_m$ and $\mathbf{z}_m$ represent the right and left singular vectors respectively. However, since $\mathbf{v}_m$ is an eigenvector of the signal autocorrelation matrix and $\mathbf{z}_m$ is an eigenvector of the autocorrelation matrix, then from the previous discussion on eigenanalysis in Section 6.1 (equation (6.24)), $\mathbf{v}_m = \mathbf{z}_m$. Similarly it can be shown that $\mathbf{u}_m = \mathbf{w}_m$. Therefore

$$\mathbf{X} = \sum_{m=0}^{p} \delta_m \mathbf{u}_m \mathbf{v}_m^H . \qquad (6.34)$$

Noting that σ_m and δ_m are the positive square roots of the eigenvalues of the signal autocorrelation matrix and autocorrelation matrix respectively (see equation (B.20) in Appendix B), then from equation (6.25)

$$\delta_m^2 = \sigma_m^2 + \eta^2 , \qquad (6.35)$$

where η^2 is the noise power. Hence equation (6.34) becomes,

$$\mathbf{X} = \sum_{m=0}^{M-1} \sqrt{\sigma_m^2 + \eta^2}\, \mathbf{u}_m \mathbf{v}_m^H + \sum_{m=M}^{p} \eta \mathbf{u}_m \mathbf{v}_m^H. \tag{6.36}$$

In this form it is apparent that the singular vectors associated with the M largest singular values of $\mathbf{X}$ (i.e. the vectors used in the left summation term) are a basis for the signal subspace, and by default, the singular vectors associated with the p-M+1 smallest singular values of $\mathbf{X}$ (i.e. the vectors used in the right summation term) form the noise subspace.

In practice, since only a limited amount of data will be available, $\mathbf{X}^H\mathbf{X}$ will only be an estimate of the true autocorrelation matrix. As a result, the previous analysis will only be approximately true.

6.3 QR FACTORIZATION

Given a matrix $\mathbf{A}$ there is a factorization such that

$$\mathbf{A} = \mathbf{QU}, \tag{6.37}$$

where $\mathbf{Q}$ is an unitary matrix (i.e. $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$) and $\mathbf{U}$ is an upper triangular matrix. Eigenanalysis and singular value decomposition methods are often performed based on iterative QR approaches to solve for the eigenvalues and eigenvectors of a matrix.

An interesting alternative to the iterative approach is to perform an approximate decomposition based on only a single QR decomposition [6-4]. For example, performing QR factorization on the signal autocorrelation matrix results in,

$$\mathbf{R}_s = \mathbf{QU}, \tag{6.41}$$

which can be expanded as

$$\begin{aligned} \mathbf{s}_0 &= u_{00}\mathbf{q}_0 \\ \mathbf{s}_1 &= u_{01}\mathbf{q}_0 + u_{11}\mathbf{q}_1 \\ &\vdots \\ \mathbf{s}_p &= u_{0p}\mathbf{q}_0 + u_{1p}\mathbf{q}_1 + \ldots + u_{pp}\mathbf{q}_p, \end{aligned} \tag{6.42}$$

where $\mathbf{s}_i$ and $\mathbf{q}_i$ represent the columns of $\mathbf{R}_s$ and $\mathbf{Q}$ respectively, and u_{ij} represents the elements of $\mathbf{U}$. If the signal matrix has rank M, then only the first M columns of $\mathbf{R}_s$ will be linearly independent. That is, the vectors $\mathbf{s}_M$, $\mathbf{s}_{M+1}$, $\mathbf{s}_{M+2}$, ..., $\mathbf{s}_p$ will be a linear combination of the first M columns of $\mathbf{R}_s$ ($\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_{M-1}$). Since the vectors $\mathbf{q}_0$, $\mathbf{q}_1$, $\mathbf{q}_2$, ..., $\mathbf{q}_p$ are orthogonal (by definition), then the columns $\mathbf{s}_M$, $\mathbf{s}_{M+1}$, $\mathbf{s}_{M+2}$, ..., $\mathbf{s}_p$ will be will also be a linear combination of the vectors $\mathbf{q}_0$, $\mathbf{q}_1$, $\mathbf{q}_2$, ..., $\mathbf{q}_{M-1}$ only, and therefore the values of $u_{ij} = 0$ for $i \geq M$. From this it is apparent that the first M columns of the matrix $\mathbf{Q}$ form an orthonormal basis set for the signal autocorrelation matrix.

Similarly the QR factorization of the autocorrelation matrix results in,

$$\mathbf{R} = \mathbf{PV}, \tag{6.43}$$

where $\mathbf{R}$ is the autocorrelation matrix, $\mathbf{P}$ is an unitary matrix and $\mathbf{V}$ is an upper triangular matrix. Again, this can be expanded as,

$$\begin{aligned} \mathbf{r}_0 &= v_{00}\mathbf{p}_0 \\ \mathbf{r}_1 &= v_{01}\mathbf{p}_0 + v_{11}\mathbf{p}_1 \\ &\vdots \\ \mathbf{r}_p &= v_{0p}\mathbf{p}_0 + v_{1p}\mathbf{p}1 + \ldots + v_{pp}\mathbf{p}_p, \end{aligned} \tag{6.44}$$

where $\mathbf{r}_i$ represents the columns of the autocorrelation matrix $\mathbf{R}$. In the case of white noise the relationship given in equation (6.13) may be used. Applying this to the set of equations (6.44) then,

$$\begin{aligned} \mathbf{s}_0 + \eta^2[1, 0, 0, \ldots, 0]^T &= v_{00}\mathbf{p}_0 \\ \mathbf{s}_1 + \eta^2[0, 1, 0, \ldots, 0]^T &= v_{01}\mathbf{p}_0 + v_{11}\mathbf{p}_1 \\ &\vdots \\ \mathbf{s}_p + \eta^2[0, 0, 0, \ldots, 1]^T &= v_{0p}\mathbf{p}_0 + v_{1p}\mathbf{p}_1 + \ldots + v_{pp}\mathbf{p}_p, \end{aligned} \tag{6.45}$$

and finally expanding $\mathbf{s}_i$ in terms of the relationship given in equation (6.45) then

$$\begin{aligned} \eta^2[1, 0, 0, \ldots, 0]^T + u_{00}\mathbf{q}_0 &= v_{00}\mathbf{p}_0 \\ \eta^2[0, 1, 0, \ldots, 0]^T + u_{01}\mathbf{q}_0 + u_{11}\mathbf{q}_1 &= v_{01}\mathbf{p}_0 + v_{11}\mathbf{p}_1 \\ &\vdots \\ \eta^2[0, 0, 0, \ldots, 1]^T + u_{0p}\mathbf{q}_0 + u_{1p}\mathbf{q}_1 + \ldots + u_{pp}\mathbf{q}_p &= v_{0p}\mathbf{p}_0 + v_{1p}\mathbf{p}_1 + \ldots + v_{pp}\mathbf{p}_p. \end{aligned} \tag{6.46}$$

By inspection the first M basis vectors of the autocorrelation matrix will not equal the first M basis vectors of the signal correlation matrix, a necessary condition if these vectors are to be used to create a signal and noise subspace. However, this problem can be overcome if the contribution of noise, η^2, is first subtracted from the main diagonal of the autocorrelation matrix before performing the QR factorization. In this case the vectors $\mathbf{q}_j$, and $\mathbf{p}_j$ will be equal.

An estimate of the noise can be made based on the empirically obtained result that

$$|v_{ii}| \simeq \frac{p+1}{p}\eta^2 \qquad \text{for } M \leq i \leq p. \tag{6.47}$$

Rearranging in terms of σ^2 and averaging over the indicated range of the index i, then

$$\hat{\eta}^2 = \frac{p}{(p+1)(p-M+1)} \sum_{i=M}^{p} |v_{ii}|. \tag{6.48}$$

Since these results are based on the asymptotic case where $p \to \infty$, the noise variance estimate will be less accurate as p decreases.

The ordering of the columns of $\mathbf{R}$ is also important. The optimum ordering is not necessarily the natural ordering resulting from the operation $\mathbf{X}^H\mathbf{X}$, but rather the ordering where the first M columns are selected to be those which are closest to being mutually orthogonal. In the case of closely spaced signals, the ordering of the columns can be determined in steps. At the first step the first column is chosen. In successive steps, the column which maximizes the minimum distance (in terms of the column indices) between it and all the other previously selected columns is chosen. The ordering of the remaining columns is not important. For example in the case of M=2, the optimum ordering is achieved by selecting the first column, then selecting the last column since it is farthest from the first column. In the case of M=3, the first two columns are selected the same way as for M=2, and the third column will be selected from a center column (e.g. for p=6 the ordering would be 0,6,2 and for p=7 the ordering would be 0,7,3 or 0,7,4).

The removal of the noise power from the main diagonal of the autocorrelation matrix, followed by the reordering of the columns of the matrix, are important additions to the QR method and result in performance almost equal to that of the more computationally demanding eigenanalysis or SVD methods.

The noise/signal subspace division is made by considering that the signals can be completely represented by the M column vectors of the matrix $\mathbf{Q}$ and consequently these vectors form the basis of the signal subspace. The remaining p-M+1 column vectors in $\mathbf{Q}$ form the basis of the noise subspace.

In practice, when only a limited amount of data is available for estimation of the autocorrelation matrix, the preceding analysis is only approximately true.

7.0 DEALING WITH THE REDUCED RANK PROBLEM

7.1 THE AUGMENTED AUTOCORRELATION MATRIX

In Section 5 a number of DF estimators were discussed (AR, ME, LP, MV, and TN) which involve computing the coefficients of an all pole filter. One common approach to computing the spectrum of each of these estimators involves the inversion of the augmented normal autocorrelation matrix ($\mathbf{R}^{-1}$). This approach assumes that the autocorrelation matrix $\mathbf{R}$ is full rank and nonsingular. In general practice this will be true since uncorrelated sensor noise (usually modelled as complex white Gaussian noise) will ensure that the autocorrelation matrix is full rank and invertible.

7.1.1 The General Solution

In Section 6 methods to estimate the signal autocorrelation matrix $\mathbf{R}_s$ were discussed. Using various decomposition methods to analyze the vector space spanned by the columns (or rows) of the autocorrelation matrix it was shown that it is possible to divide this vector space into a signal subspace and noise subspace. Using the matrix formed from the signal subspace as the signal autocorrelation matrix estimate, the idea was to enhance all pole estimator performance by removing the noise before estimating the filter coefficients. The first difficulty with this approach, however, is that the signal autocorrelation matrix (and its estimate) has rank M, so if $M \leq p$, this matrix is not invertible.

In reality the difficulty arises since $p+1$ coefficients are being used to determine M signal bearings. In other words there are more coefficients than necessary (assuming $M \leq p$) and as a result an infinite number of solutions are possible. This difficulty can be overcome by examining the all pole estimators in more detail. A common feature of all these methods is the minimization of the filter output variance σ^2 given by the equation

$$\sigma^2 = \mathbf{a}^H \mathbf{R} \mathbf{a}, \tag{7.1}$$

subject to some constraint (dependent on the estimator used) on the choice of the filter coefficient vector $\mathbf{a}$ (the vector from which the DF spectrum is generated).

To incorporate the constraint into the function to be minimized, the Lagrange multiplier technique can be used. Based on this technique a new function is defined such that

$$F = \mathbf{a}^H \mathbf{R} \mathbf{a} + \delta c(\mathbf{a}), \tag{7.2}$$

where the function $c(\mathbf{a}) = 0$ when the constraint on $\mathbf{a}$ is satisfied. Minimizing F with respect to $\mathbf{a}^H$ results in an equation of the form,

$$\mathbf{R}\mathbf{a} + \delta \mathbf{c} = 0, \tag{7.3}$$

where the constraint vector $\mathbf{c}$ is defined as

$$\mathbf{c} = \begin{bmatrix} \frac{\partial c(\mathbf{a})}{\partial a_0^*} \\ \frac{\partial c(\mathbf{a})}{\partial a_1^*} \\ \vdots \\ \frac{\partial c(\mathbf{a})}{\partial a_p^*} \end{bmatrix}. \tag{7.4}$$

Minimizing F with respect to δ results in the original constraint given by,

$$c(\mathbf{a}) = 0. \tag{7.5}$$

The solution of the coefficient vector **a** can then be determined by solving equations (7.3) and (7.5).

A feature of the derivation of the solution coefficient vector **a** is that it involves the term **Ra** (see equation (7.3)). If **R** is invertible, **a** is easily isolated by premultiplying the term by $\mathbf{R}^{-1}$. In the case where **R** is not full rank, then determination of **a** must be handled differently.

The effect of the constraint on the minimization of equation (7.2) is to reduce the number of degrees of freedom in the choice of the coefficient vector **a** by 1. In otherwords, if **R** is a $(p+1)\times(p+1)$ autocorrelation matrix, then the choice of the vector **a** will have p degrees of freedom. The optimum choice for the vector **a** is the one wich results in the lowest variance; ideally $\sigma^2 = 0$. However, if the autocorrelation matrix is full rank, the loss of a degree of freedom means that it is impossible to choose a vector **a** which simultaneously satisfies the given constraint and sets the output filter variance to zero. This would require a vector with $p+1$ degrees of freedom.

In the case where the estimated signal autocorrelation matrix $\mathbf{R}_s$ is used in place of **R**, the situation changes since $\mathbf{R}_s$ is not full rank (assuming $M \leq p$) but has rank M. Since this reduces the number of degrees of freedom required from $p+1$ to M, the vector **a** can be chosen to satisfy both the constraint and the relationship

$$\mathbf{a}^H\mathbf{R}_s\mathbf{a} = 0. \tag{7.6}$$

Expanding $\mathbf{R}_s$ in terms of the signal data matrix $\mathbf{X}_s$, equation (7.6) can be rewritten as,

$$\mathbf{a}^H\mathbf{X}_s\mathbf{X}_s^H\mathbf{a} = \left|\mathbf{a}^H\mathbf{X}_s\right|^2 = 0. \tag{7.7}$$

This equation may only be satisfied if

$$\mathbf{X}_s^H\mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7.8}$$

Premultiplying this result by $\mathbf{X}_s$ then,

$$\mathbf{X}_s\mathbf{X}_s^H\mathbf{a} = \mathbf{R}_s\mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7.9}$$

Assuming that $M \leq p$, then there are an infinite number of solutions to this equation. The family of all possible solutions may be determined by examining the eigenstructure of $\mathbf{R}_s$. Using the notation developed in Section 5.1, then,

$$\mathbf{R}_s = \sum_{m=0}^{p} \lambda_m \mathbf{v}_m \mathbf{v}_m^H, \tag{7.10}$$

where λ_m represents the eigenvalues of $\mathbf{R}$ (and $\mathbf{R}_s$) ordered so that $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_p$ and $\mathbf{v}_m$ represents the corresponding eigenvectors of $\mathbf{R}$ (and $\mathbf{R}_s$). Additionally $\mathbf{R}_s$ has rank $M \leq p$, so that $\lambda_M = \lambda_{M+1} = \ldots = \lambda_p = 0$.

Since the eigenvectors form an orthonormal basis set ($\mathbf{v}_i^H\mathbf{v}_j = 0$ for $i \neq j$, and $\mathbf{v}_i^H\mathbf{v}_i = 1$) and $\mathbf{R}_s$ is formed from the first M eigenvectors, then the vector $\mathbf{a}$ must be orthogonal to these M eigenvectors to satisfy equation (7.9). Thus the vector $\mathbf{a}$ is a linear combination of the last p-M+1 eigenvectors ($\mathbf{v}_M$, $\mathbf{v}_{M+1}$, $\mathbf{v}_{M+2}$, ..., $\mathbf{v}_p$). One possible representation is given by

$$\mathbf{a} = \left(\sum_{m=M}^{p} \delta_m \mathbf{v}_m \mathbf{v}_m^H \right)\mathbf{q}, \tag{7.11}$$

where the values of δ_m are arbitrarily chosen, and the vector $\mathbf{q}$ is chosen so that $\mathbf{a}$ satisfies the constraint condition in equation (7.5). From equation (7.11) a new matrix can be defined as

$$\mathbf{R}_s^\Upsilon = \sum_{m=M}^{p} \delta_m \mathbf{v}_m \mathbf{v}_m^H, \tag{7.12}$$

where the symbol Υ is defined here as the *eigeninverse* operator which sets all the nonzero eigenvalues of the matrix $\mathbf{R}_s$ to 0 and all the zero eigenvalues to some arbitrary value. The choice of the new eigenvalues will be discussed later on in Sections 7.1.2.1 to 7.1.2.3. Note that $\mathbf{R}_s^\Upsilon$ is formed from the eigenvectors of the noise subspace of $\mathbf{R}$. Using this new representation, equation (7.11) becomes

$$\mathbf{a} = \mathbf{R}_s^\Upsilon \mathbf{q}. \tag{7.13}$$

The selection of a suitable constraint vector $\mathbf{q}$ and subsequent derivation of $\mathbf{a}$ may be simplified considerably by noting that when the autocorrelation matrix has full rank, equation (7.3) can be rewritten as

$$\mathbf{a} = -\mathbf{R}^{-1}\delta\mathbf{c}. \tag{7.14}$$

By setting $\mathbf{q} = -\delta\mathbf{c}$ in equation (7.13) the solution of the coefficient vector $\mathbf{a}$ can be determined by solving,

$$\mathbf{a} = -\mathbf{R}_s^{\mathsf{T}}\delta\mathbf{c}, \tag{7.15}$$

subject to the constraint

$$c(\mathbf{a}) = 0. \tag{7.16}$$

Solving these equations is identical to the solution procedure for the full rank case (involving equations (7.3) or (7.14), and (7.5)), except that the matrix $\mathbf{R}_s^{\mathsf{T}}$ has been exchanged for $\mathbf{R}^{-1}$. In other words, the noise removal techniques discussed in Section 6 can be used to enhance any of the all pole DF estimators described previously in Section 5 simply by replacing $\mathbf{R}^{-1}$ by the matrix $\mathbf{R}_s^{\mathsf{T}}$. For example, the enhanced version of the MV method (see equation (5.64)) is given by

$$S(\phi) = \frac{1}{\mathbf{e}^{\mathsf{H}}\mathbf{R}_s^{\mathsf{T}}\mathbf{e}}. \tag{7.17}$$

7.1.2 The Eigeninverse Solution

Up to this point no criteria has been given for the selection of the eigenvalues of the eigeninverse solution $\mathbf{R}_s^{\mathsf{T}}$. The choice of these eigenvalues has an effect on the size and location of spurious peaks in the DF spectrum, and in the case where the estimated autocorrelation matrix is used, it also affects the location of peaks corresponding to actual signal bearings as well. As a result, the choice of the eigenvalues can have a significant effect on the accuracy of the DF estimator.

Three different approaches are examined in the next three sections.

7.1.2.1 The White Noise Approach

If the eigenvalues of the eigeninverse solution $\mathbf{R}_s^{\mathsf{T}}$ are chosen, the resulting DF spectrum will have M peaks corresponding to the bearings of the M actual signals, and up to p-M spurious peaks with arbitrary locations. A useful solution is one that uniformly suppresses these spurious peaks or whitens the spectrum. This reduces the chances of confusing a spurious peak with an actual signal peak.

Since the eigenvalues of a white noise process are all equal, then the estimated spectrum can be whitened by setting the nonzero eigenvalues of $\mathbf{R}_s^{\mathsf{T}}$ to 1 [7-1]. Defining this in this report as the *whitened eigeninverse*, it is expressed in terms of equation (7.12) as,

$$\mathbf{R}_s^{\mathsf{T}} = \sum_{m=M}^{p} \mathbf{v}_m \mathbf{v}_m^{\mathsf{H}}. \tag{7.18}$$

An equivalent definition using the Moore-Penrose pseudoinverse [7-2] is given by

$$\mathbf{R}_s^{\mathsf{T}} = \mathbf{I} - \mathbf{R}_s\mathbf{R}_s^{\#}, \tag{7.19}$$

where the # is used to represent the pseudoinverse operation. The pseudoinverse of a

matrix can be defined in terms of its singular value decomposition. Using an arbtrary matrix $\mathbf{Q}$ as an example, then

$$\mathbf{Q}^{\#} = \sum_{m=0}^{K} \frac{1}{\sigma_m} \mathbf{u}_m \mathbf{v}_m^{H} \tag{7.20}$$

where K is the rank of $\mathbf{Q}$ (and is less than or equal to the smallest dimension of $\mathbf{Q}$), σ_m represents the singular values of $\mathbf{Q}$, and $\mathbf{u}_m$ and $\mathbf{v}_m$ represent the left and right singular vectors of $\mathbf{Q}$ respectively.

Alternatively, under the special condition where the signal data matrix has dimensions $p \times M$, the whitened eigeninverse can also be defined as,

$$\mathbf{R}_s^{\dagger} = \mathbf{I} - \mathbf{X}_s \left(\mathbf{X}_s^{H} \mathbf{X}_s\right)^{-1} \mathbf{X}_s^{H} \tag{7.21}$$

7.1.2.2 Approach of Johnson and Degraff

A second approach proposed by Johnson and Degraff [7-3] to the selection of eigenvalues is based on the solution which results when the autocorrelation matrix is inverted first, and the signal/noise subspace division performed afterwards. The result of the matrix inversion is given by,

$$\mathbf{R}^{-1} = \sum_{m=0}^{M-1} \frac{1}{\lambda_m} \mathbf{v}_m \mathbf{v}_m^{H} + \sum_{m=M}^{p} \frac{1}{\lambda_m} \mathbf{v}_m \mathbf{v}_m^{H} \tag{7.22}$$

where the first term represents the signal subspace, and the second term represents the noise subspace. From the analysis given previously, the enhanced form of $\mathbf{R}^{-1}$ lies entirely in the noise subspace, in which case the signal subspace eigenvalues are set to 0. That is

$$\mathbf{R}_s^{\dagger} = \sum_{m=M}^{p} \frac{1}{\lambda_m} \mathbf{v}_m \mathbf{v}_m^{H}. \tag{7.23}$$

7.1.2.3 Approach of Wax and Kailath

The third approach proposed by Wax and Kailath [7-4] is based on an examination of the maximum likelihood solution to the bearing estimation problem for a single signal. The maximum likelihood method seeks to find the signal which best fits the data. For a single signal, assuming zero mean Gaussian noise, the maximum likelihood solution reduces to a least squares minimization which is given by the following,

$$\epsilon^2 = \sum_{i=1}^{T} \left| \mathbf{x}(i) - c(i)\mathbf{e} \right|^2, \tag{7.24}$$

where the minimization of ϵ^2 is performed with respect to $c(i)$ and ϕ, the vector $\mathbf{x}(i)$ represents a single sample of the sensor data measured at the time instance represented by i, $c(i)$ represents the phase and amplitude of the signal measured at sensor 0, and $\mathbf{e}$ is the steering vector (i.e. it determines the signal phase and amplitude at each of the other sensors as a function of ϕ relative to sensor 0).

Minimizing equation (7.24) with respect to $c(i)$ results in the expression

$$c(i) = \frac{\mathbf{e}^H \mathbf{x}(i)}{\mathbf{e}^H \mathbf{e}}. \tag{7.25}$$

Replacing $c(i)$ by this expression and noting that $\mathbf{e}^H\mathbf{e} = N$, then the minimization of ϵ^2 can be expressed as a function of ϕ only. That is, equation (7.24) becomes,

$$\epsilon^2 = \sum_{i=1}^{T} \left| \mathbf{x}(i) - \frac{1}{N} \mathbf{e}\mathbf{e}^H \mathbf{x}(i) \right|^2, \tag{7.26}$$

which can be expanded as,

$$\epsilon^2 = \sum_{i=1}^{T} \mathbf{x}^H(i)\mathbf{x}(i) \; - \; (\frac{2}{N} - \frac{1}{N^2}) \sum_{i=1}^{T} \mathbf{x}^H(i)\mathbf{e}\mathbf{e}^H\mathbf{x}(i). \tag{7.27}$$

Given that the above expression is real valued and greater than or equal to zero, and since the first term is not a function of ϕ, then minimizing this expression is equivalent to maximizing the second term. Ignoring the constant multiplier term, the maximum likelihood solution can be found by maximizing,

$$\sum_{i=1}^{T} \mathbf{x}^H(i)\mathbf{e}\mathbf{e}^H\mathbf{x}(i). \tag{7.28}$$

Additionally since $\mathbf{e}^H\mathbf{x}(i)$ is a scalar value, the above expression can be rewritten as,

$$\sum_{i=1}^{T} \mathbf{e}^H\mathbf{x}(i)\mathbf{x}^H(i)\mathbf{e}. \tag{7.29}$$

The covariance estimate of the autocorrelation matrix for multiple samples is defined by,

$$\hat{\mathbf{R}} = \frac{1}{T} \sum_{i=1}^{T} \mathbf{x}(i)\mathbf{x}^H(i). \tag{7.30}$$

Using this definition for $\hat{\mathbf{R}}$, the maximum likelihood solution can be determined by maximizing the function,

$$\mathbf{e}^H\hat{\mathbf{R}}\mathbf{e}, \tag{7.31}$$

which is in fact the classical spectral estimator. In terms of the eigenvalues and eigenvectors of the covariance estimate of $\mathbf{R}$, this equation can be reformulated as,

$$\mathbf{e}^H\Big(\lambda_0\mathbf{v}_0\mathbf{v}_0^H + \sum_{m=1}^{N-1}\lambda_m\mathbf{v}_m\mathbf{v}_m^H\Big)\mathbf{e}. \tag{7.32}$$

Using the fact that,

$$\sum_{m=0}^{N-1}\mathbf{v}_m\mathbf{v}_m^H = \mathbf{I}, \tag{7.33}$$

and expressing $\mathbf{v}_0\mathbf{v}_0^H$ in terms of the other eigenvectors, then equation (7.32) becomes,

$$\mathbf{e}^H\Big(\lambda_0\mathbf{I} - \sum_{m=1}^{N-1}(\lambda_0 - \lambda_m)\mathbf{v}_m\mathbf{v}_m^H\Big)\mathbf{e}. \tag{7.34}$$

Since the term,

$$\mathbf{e}^H\lambda_0\mathbf{I}\mathbf{e} = \lambda_0\mathbf{e}^H\mathbf{e} = \frac{\lambda_0}{N}, \tag{7.35}$$

is constant with respect to ϕ, its effect can be ignored. This gives,

$$-\mathbf{e}^H\Big(\sum_{m=1}^{N-1}(\lambda_0 - \lambda_m)\mathbf{v}_m\mathbf{v}_m^H\Big)\mathbf{e}. \tag{7.36}$$

Finally maximizing this last expression is equivalent to minimizing the negative reciprocal given by,

$$\frac{1}{\mathbf{e}^H\Big(\sum_{m=1}^{N-1}(\lambda_0 - \lambda_m)\mathbf{v}_m\mathbf{v}_m^H\Big)\mathbf{e}}. \tag{7.37}$$

Inspection of equation (7.37) reveals that it is in fact an enhanced form of the MV method for a single signal, and that the summation term in the brackets is a form of the matrix $\mathbf{R}_s^I$. Using this form of $\mathbf{R}_s^I$, Wax and Kailath proposed an extension to the multiple signal case (applying the previous analysis to determine an exact solution for the multiple signal case cannot be done in any simple fashion) which is defined here as,

$$\mathbf{R}_s^I = \sum_{m=M}^{p}(\bar{\lambda}_s - \lambda_m)\mathbf{v}_m\mathbf{v}_m^H, \tag{7.38}$$

where $\bar{\lambda}_s$ serves the same function as λ_0 in equation (7.38) and is defined as the mean of the signal eigenvalues given by,

$$\bar{\lambda}_s = \frac{1}{M}\sum_{m=0}^{M-1}\lambda_m. \tag{7.39}$$

7.2 THE NORMAL AUTOCORRELATION MATRIX

Throughout most of this report, autoregressive DF estimators based on the augmented normal equation formulation of the autocorrelation matrix have been discussed. However, an alternate formulation using only the normal equations was given, and is repeated here as (see equation (5.11) in Section 5.2.1),

$$\mathbf{R}_p\mathbf{w} = -\mathbf{r}, \tag{7.40}$$

where $\mathbf{R}_p$ is a $p \times p$ autocorrelation matrix, $\mathbf{w}$ is the tap weight coefficient vector (from which the DF spectrum is derived), and $\mathbf{r}$ was defined previously in equations (5.13) and (2.5).

The signal/noise subspace technique can be used to enhance $\mathbf{R}_p$ in the same manner as in the augmented autocorrelation case simply by setting the smallest p-M eigenvalues of $\mathbf{R}_p$ to zero. The result is given by

$$\mathbf{R}_s = \sum_{m=0}^{M-1} \lambda_m \mathbf{v}_m \mathbf{v}_m^H, \tag{7.41}$$

where in this case $\mathbf{R}_s$ is the signal autocorrelation matrix estimate of $\mathbf{R}_p$.

In equation (7.40) the quantity of interest is $\mathbf{w}$, and it is normally calculated as

$$\mathbf{w} = -\mathbf{R}_p^{-1}\mathbf{r}. \tag{7.42}$$

However, if $\mathbf{R}_p$ is replaced by $\mathbf{R}_s$ this solution is no longer applicable since $\mathbf{R}_s$ is not full rank and therefore not invertible. However, analysis of the equation

$$\mathbf{R}_s\mathbf{w} = -\mathbf{r}, \tag{7.43}$$

does yield some insight into the solution to the problem. For example $\mathbf{w}$ can be represented as a linear combination of the eigenvectors of the autocorrelation matrix $\mathbf{R}_p$. One possible representation is given by,

$$\mathbf{w} = \left(\sum_{m=0}^{M-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H + \sum_{m=M}^{p-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H \right) \mathbf{q}, \tag{7.44}$$

where in this case $\mathbf{q}$ is a column vector of ones and the values of δ_m are chosen so that $\mathbf{w}$ is a solution to equation (7.42). The above representation also clearly distinguishes between the signal subspace eigenvectors (first summation term), and the noise subspace eigenvectors (second summation term) of the autocorrelation matrix.

Using this last representation for $\mathbf{w}$, and the representation of $\mathbf{R}_s$ given by equation (7.41), equation (7.43) can be expanded as,

$$\mathbf{R}_s\mathbf{w} = \left(\sum_{m=0}^{M-1} \lambda_m \mathbf{v}_m \mathbf{v}_m^H \right) \left(\sum_{m=0}^{M-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H + \sum_{m=M}^{p-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H \right) \mathbf{q} = -\mathbf{r}. \tag{7.45}$$

Simplifying, by taking advantage of the fact that the eigenvectors form an orthonormal set, the result can be expressed as

$$\mathbf{R}_s\mathbf{w} = \left(\sum_{m=0}^{M-1} \lambda_m \delta_m \mathbf{v}_m \mathbf{v}_m^H \right) \mathbf{q} = -\mathbf{r}. \tag{7.46}$$

From the form of this equation, it is clear that the choice of values of δ_m for $M \leq m \leq p$ has no effect on the solution of equation (7.43). In other words, these values of δ_m may be arbitrarily chosen and equation (7.43) will still be satisfied.

Based on this observation it is possible to derive the general solution to equation (7.43) for $\mathbf{w}$. The tap weight coefficient vector $\mathbf{w}$ can be represented as

$$\mathbf{w} = \mathbf{w}_s + \mathbf{w}_n, \tag{7.47}$$

where $\mathbf{w}_s$ is the part of $\mathbf{w}$ which is formed from the signal subspace of $\mathbf{R}_p$ in equation (7.46) and is defined (based on equation (7.44)) as,

$$\mathbf{w}_s = \left(\sum_{m=0}^{M-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H \right) \mathbf{q}, \tag{7.48}$$

and $\mathbf{w}_n$ is the part of $\mathbf{w}$ which is formed from the noise subspace of $\mathbf{R}_p$ and is defined as,

$$\mathbf{w}_n = \left(\sum_{m=M}^{p-1} \delta_m \mathbf{v}_m \mathbf{v}_m^H \right) \mathbf{q}. \tag{7.49}$$

As stated before the choice of δ_m for $\mathbf{w}_n$ is arbitrary so that equation (7.49) represents the general solution for $\mathbf{w}_n$ and need not be simplified any further. The choice of δ_m for $\mathbf{w}_s$, on the other hand, is critical and must be chosen so that equation (7.43) is satisfied.

To solve for $\mathbf{w}_s$, equation (7.43) is rewritten as

$$\mathbf{R}_s\mathbf{w} = \mathbf{R}_s\mathbf{w}_s = -\mathbf{r}. \tag{7.50}$$

Since $\mathbf{R}_s$ is not invertible, a new matrix, derived from the noise subspace eigenvectors (but not eigenvalues) of $\mathbf{R}_p$ is chosen to act as a "catalyst" in the solution. Namely,

$$\mathbf{R}_n = \sum_{m=M}^{p-1} \mathbf{v}_m \mathbf{v}_m^H. \tag{7.51}$$

This matrix is formed from the noise subspace of $\mathbf{R}$ and has the properties that,

$$\mathbf{R}_n\mathbf{w}_s = \underline{0}, \tag{7.52}$$

and

$$\mathbf{R}_n\mathbf{r} = \underline{0} \tag{7.53}$$

(this second property is a result of the fact that based on equation (7.46) $\mathbf{r}$ is a linear combination of the signal subspace eigenvectors which are orthogonal to the eigenvectors of $\mathbf{R}_n$). Using the first property (equation (7.52)), equation (7.50) can be rewritten as

$$(\mathbf{R}_s + \mathbf{R}_n)\mathbf{w}_s = -\mathbf{r}. \tag{7.54}$$

The matrix $(\mathbf{R}_s + \mathbf{R}_n)$ is invertible, so the solution for $\mathbf{w}_s$ is given by,

$$\mathbf{w}_s = -(\mathbf{R}_s + \mathbf{R}_n)^{-1}\mathbf{r}. \tag{7.55}$$

Again using the eigenvector representation, equation (7.55) can be expanded as

$$\mathbf{w}_s = -\left(\sum_{m=0}^{M}\frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H + \sum_{m=M}^{p-1}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{r}. \tag{7.56}$$

Recognizing the second summation term as $\mathbf{R}_n$, and taking advantage of the relationship expressed in equation (7.53), then the solution for $\mathbf{w}_s$ simplifies,

$$\mathbf{w}_s = -\left(\sum_{m=0}^{M}\frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{r}. \tag{7.57}$$

Therefore, the general solution for $\mathbf{w}$ in terms of the eigenvalues and eigenvectors of the autocorrelation matrix is given by,

$$\mathbf{w} = -\left(\sum_{m=0}^{M}\frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{r} + \left(\sum_{m=M}^{p-1}\delta_m\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{q}. \tag{7.58}$$

An infinite number of solutions exist for $\mathbf{w}$ since the values of δ_m can be arbitrarily chosen.

7.2.1 The Pseudoinverse Solution

In the preceding discussion regarding the general solution of the tap weight coefficient vector $\mathbf{w}$, an infinite number of solutions were shown to exist. When the true autocorrelation matrix $\mathbf{R}_p$ is used, each solution will result in a DF spectrum with M peaks corresponding to the bearing of the M actual signals, and up to p-M spurious peaks with arbitrary location. The most useful solution is the one that suppresses these peaks.

In the general solution of $\mathbf{w}$ expressed by equation (7.58), the first summation term represents the contribution of the signal subspace eigenvectors of the autocorrelation matrix $\mathbf{R}_p$ and remains invariant for all possible solutions of $\mathbf{w}$. The second term represents the variable part of $\mathbf{w}$. From this it can be concluded that the size and location of the signal peaks in the DF spectrum are controlled by the first term, and the spurious peaks are controlled by the second term. Consequently the size of the spurious peaks in the spectrum can be minimized by minimizing the second summation term. Since this corresponds to setting $\delta_m = 0$, the desired form of $\mathbf{w}$ is given by [7-5],

$$\mathbf{w} = -\left(\sum_{m=0}^{M}\frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{r} \tag{7.59}$$

The expression inside the brackets is the pseudoinverse (previously defined by equation (7.20)) of $\mathbf{R}_s$ and consequently equation (7.58) can also be written as,

$$\mathbf{w} = -\mathbf{R}_s^{\#}\,\mathbf{r}, \tag{7.60}$$

where the symbol # indicates the pseudoinverse operation.

This solution for $\mathbf{w}$ is also called the minimum norm solution since this solution of $\mathbf{w}$ has the minimum norm of all possible solutions where the norm of $\mathbf{w}$ is given by

$$|\mathbf{w}| = \sqrt{\mathbf{w}^H\mathbf{w}}. \tag{7.61}$$

This can be shown by expanding $\mathbf{w}$ using the relationship expressed in equation (7.58), and squaring to get,

$$|\mathbf{w}|^2 = \left[\mathbf{r}^H \sum_{m=0}^{M} \frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H + \mathbf{q}^H \sum_{m=M}^{p-1} \delta_m\mathbf{v}_m\mathbf{v}_m^H\right]\left[\left(\sum_{m=0}^{M} \frac{1}{\lambda_m}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{r} + \left(\sum_{m=M}^{p-1} \delta_m\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{q}\right]. \tag{7.62}$$

By taking advantage of the fact that the eigenvectors form an orthonormal set, this expression simplifies to

$$|\mathbf{w}|^2 = \mathbf{r}^H\mathbf{r} \sum_{m=0}^{M} \frac{1}{\lambda_m^2} + \mathbf{q}^H\mathbf{q} \sum_{m=M}^{p-1} \delta_m^2. \tag{7.63}$$

By inspection it is clear that $|\mathbf{w}|^2$ and correspondingly the norm $|\mathbf{w}|$ are minimized when $\delta_m = 0$. Therefore equation (7.59) represents the minimum norm solution of $\mathbf{w}$.

8.0 ENHANCED ALL POLE ESTIMATORS

In the previous two sections the techniques to remove a significant portion of the noise from the data or autocorrelation matrix (Section 6) and generate a suitable replacement for the inverse autocorrelation matrix were discussed (Section 7). Of particular interest is the improvement that results in all pole DF estimators when these techniques are used. Enhancing these estimators in this manner is equivalent to using an ARMA model without the need to explicitly compute the MA filter coefficients (since these coefficients ideally only model the noise and are therefore not required to generate the DF spectrum). As was discussed in Section 4, the ARMA filter provides the best model of the sensor information in a signal plus noise environment.

MA estimators can also be improved using the enhancements discussed in Section 6. However, since for small tactical sensor arrays MA estimators are generally resolution limited by the number and spacing of sensors, and not the noise, these enhancements are not discussed here.

Enhancements to ARMA estimators have been proposed [5-1], but since these methods generally involve a computationally extensive nonlinear search procedure, they are not discussed in this report.

From the discussion in Section 7 it was shown that an infinite number of solutions exist as possible replacements for the inverse autocorrelation matrix which is a central feature of the all pole estimators discussed in Section 5. In particular the eigeninverse, the weighted eigeninverse, and pseudoinverse solutions have all been proposed as useful solutions. The result is numerous possible enhanced estimators. To limit discussion on these enhanced estimators, only those which have appeared in the open literature are discussed. These modified estimators are also divided into two categories: linear prediction, and Capon.

8.1 ENHANCED LINEAR PREDICTION ESTIMATORS

The linear prediction estimators discussed in Section 5 are considered to include the Autoregressive, Maximum Entropy and all linear prediction methods. In the following two sections two enhanced linear prediction estimators are described. The first is based on using the augmented normal equation formulation, and the second is based on using just the normal equation formulation.

8.1.1 Minimum Norm Method

In the Minimum Norm (MNorm) method [8-1] the linear prediction equation given by,

$$S_{LP}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}^{-1}\mathbf{u}\mathbf{u}^H\mathbf{R}^{-1}\mathbf{e}}, \tag{8.1}$$

is modified to become

$$S_{MNorm}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}_s^I\mathbf{u}\mathbf{u}^H\mathbf{R}_s^I\mathbf{e}}, \tag{8.2}$$

where the inverse autocorrelation matrix is replaced by the whitened eigeninverse matrix (equation (7.18)). In terms of the eigenvalues and eigenvectors of the autocorrelation matrix, the above expression can be rewritten as,

$$S_{MNorm}(\phi) = \frac{1}{\mathbf{e}^H\left(\sum_{m=M}^{p} \mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{u}\mathbf{u}^H\left(\sum_{m=M}^{p} \mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{e}}. \tag{8.3}$$

The name Minimum Norm comes from the fact that the filter coefficient vector given by,

$$\mathbf{a} = \sigma^2\mathbf{R}_s^T\mathbf{u}, \tag{8.4}$$

has the minimum norm property. That is, although there are an infinite number of solutions for the vector **a** which satisfy the equation

$$\mathbf{R}_s\mathbf{a} = \sigma^2\mathbf{u}, \tag{8.5}$$

(where $\mathbf{R}_s$ is the estimated signal autocorrelation matrix defined in Section 6), the solution selected is the one which minimizes the vector norm $|\mathbf{a}^H\mathbf{a}|$ subject to the constraint $a_0 = 1$.

8.1.2 Modified Forward-Backward Linear Prediction

Like the Minimum Norm method the Modified Forward-Backward Linear Prediction (MFBLP) method [8-2] is based on enhancing the forward backward linear prediction method [5-5]. In the MFBLP method, the normal equations defined by

$$\mathbf{R}_p\mathbf{w} = -\mathbf{r}, \tag{8.6}$$

are used, not the augmented normal equations. The enhanced solution for **w** is given by,

$$\mathbf{w} = -\mathbf{R}_s^{\#}\mathbf{r}, \tag{8.7}$$

where in this case $\mathbf{R}_s^{\#}$ is the pseudoinverse (see equation (7.20)) of the signal autocorrelation matrix derived from $\mathbf{R}_p$.

Once the vector **w** has been estimated, the vector **a** is determined using

$$\mathbf{a} = \begin{bmatrix} 1 \\ \cdots \\ \mathbf{w} \end{bmatrix}, \tag{8.8}$$

and the DF spectrum is computed as,

$$S_{MFBLP}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{a}\mathbf{a}^H\mathbf{e}}. \tag{8.9}$$

Another similarity to the Minimum Norm method is that the solution of the vector **a** determined by the MFBLP method has the minimum norm property. That is, although there are an infinite number of solutions for the vector **w** which satisfy the equation

$$\mathbf{R}_s\mathbf{w} = -\mathbf{r}, \tag{8.10}$$

the solution selected is the one which minimizes the vector norm $|\mathbf{w}^H\mathbf{w}|$. If the vector $\mathbf{w}$ is a minimum norm solution, then so will the vector $\mathbf{a}$. From this it would seem that the Minimum Norm method and the MFBLP method are identical. However, the difference between the two methods lies in the signal/noise subspace division. For the MFBLP method this division is performed on the normal autocorrelation matrix $\mathbf{R}_p$ and for the Minimum Norm method this division is performed on the larger augmented autocorrelation matrix $\mathbf{R}$.

8.2 ENHANCED CAPON ESTIMATORS

The major difference between the Linear Prediction estimators and the Capon estimators (which include the Minimum Variance and Thermal Noise methods) is the manner in which the all pole filter coefficients are determined from the inverse autocorrelation matrix. Linear Prediction estimators select a single column from this matrix while Capon estimators use a linear combination of these columns.

Two enhanced Capon estimators are described in the following sections.

8.2.1 MUSIC Method

The Multiple Signal Classification (MUSIC) method [7-1] can be described as an enhanced version of either the Minimum Variance or the Thermal Noise methods. In either case the inverse autocorrelation matrix is replaced by the whitened eigeninverse matrix (equation (7.18)). The equivalence of the two enhancements (enhanced Minimum Variance and enhanced Thermal Noise) is a result of the fact that the eigenvalues of the replacement matrix are set to unity so that

$$\mathbf{R}_s^I = (\mathbf{R}_s^I)^2. \tag{8.11}$$

The resultant DF estimator has the form,

$$S_{MUSIC}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}_s^I\mathbf{e}}. \tag{8.12}$$

In terms of the eigenvalues and eigenvectors of the autocorrelation matrix, the above expression can be rewritten as,

$$S_{MUSIC}(\phi) = \frac{1}{\mathbf{e}^H\left(\sum_{m=M}^{p}\mathbf{v}_m\mathbf{v}_m^H\right)\mathbf{e}}. \tag{8.13}$$

8.2.2 Eigenvector Method

The Eigenvector (EV) method [7-3] can also be represented by the MUSIC estimator equation (8.12), that is,

$$S_{EV}(\phi) = \frac{1}{\mathbf{e}^H\mathbf{R}_s^I\mathbf{e}}. \tag{8.14}$$

The difference between the two methods is in the definition of $\mathbf{R}_s^I$. In the EV method $\mathbf{R}_s^I$ is the eigeninverse matrix based on the approach by Johnson and DeGraff (equation (7.23)). Using this definition, equation (8.11) is no longer true, so that the EV method can be viewed as an enhancement of the Minimum Variance estimator, but not the Thermal Noise estimator.

In terms of the eigenvalues and eigenvectors of the autocorrelation matrix, the Minimum Variance DF spectrum can also be defined as,

$$S_{EV}(\phi) = \frac{1}{\mathbf{e}^H\left(\sum_{m=M}^{p} \frac{1}{\lambda_m} \mathbf{v}_m \mathbf{v}_m^H\right)\mathbf{e}}. \tag{8.15}$$

8.3 A COMPARISON OF ENHANCED DF ESTIMATORS

In this section, two basic enhanced estimators plus variants are compared. The enhanced estimators can be summarized as:

1. Enhanced Autoregressive:
   Enhanced Maximum Entropy:
   Enhanced Linear Prediction:

$$S_{ELP}(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}_s^I \mathbf{u}\mathbf{u}^H \mathbf{R}_s^I \mathbf{e}} \tag{8.16}$$

2. Enhanced Minimum Variance:

$$S_{EMV}(\phi) = \frac{1}{\mathbf{e}^H \mathbf{R}_s^I \mathbf{e}} \tag{8.17}$$

Note that the first category includes the Minimum Norm and Modified Forward Backward Linear Prediction methods discussed in Sections 8.1.1 and 8.1.2 respectively, and the second category includes the MUSIC and the Eigenvector methods discussed in Sections 8.2.1 and 8.2.2 respectively.

Variants include the different approaches to determining the eigeninverse, namely, the whitened eigeninverse, the Johnson and Degraff approach, and the Wax and Kailath approach, as well as the approximate whitened eigeninverse solution based on QR Factorization. For comparison purposes, the whitened eigeninverse based estimators, namely the Minimum Norm method and the MUSIC method, are adopted here as the basic enhanced estimators.

The simulations that follow are also based on the identical data used previously for comparing DF estimators in Section 5.4. The optimum model order p for spatial smoothing purposes was based on a comparison of enhanced estimator accuracy versus the model order, the results of which are shown in Figure 8.1. The variance of the bearing errors for each value of p was computed from 300 trials at a fixed signal to noise ratio of 50 dB. With the exception of the variants using the approach of Johnson and Degraff, the model order chosen was $p = 5$. For those variants using the approach of Johnson and Degraff, the chosen model order was $p = 4$.

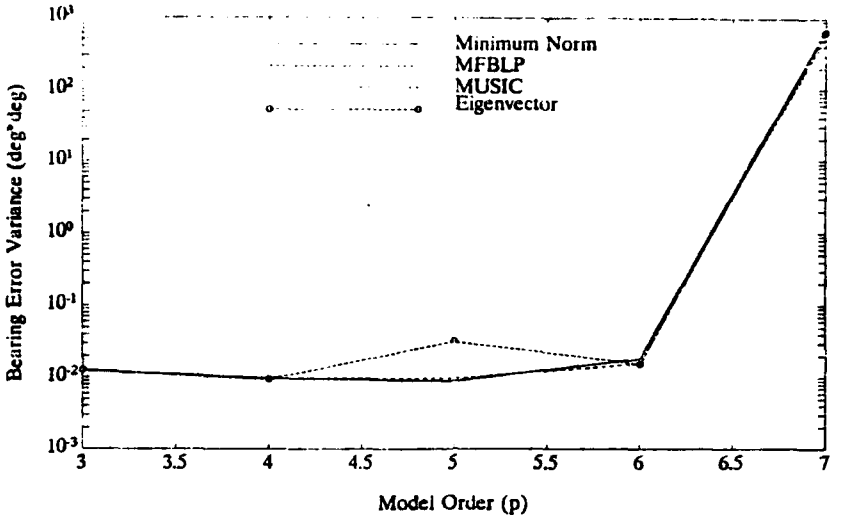


FIGURE 8.1: Bearing error variance as a function of the model order p

Figure 8.2 compares the performance of the Linear Prediction method versus the Minimum Norm and Modified Forward-Backward Linear Prediction methods, while Figure 8.3 compares the performance of the Minimum Variance method with the MUSIC method. Threshold performance for all these methods is dominated by the problem of the peaks in the estimated DF spectrum corresponding to the 40 and 50 degree bearing signals beginning to merge into a single peak. In both cases the enhanced methods outperformed the unenhanced methods in terms of accuracy and threshold performance. Interestingly, in comparing these results to Figure 5.4, the Adaptive Angular Response method still has the best threshold performance (but not accuracy) of any of the methods described so far.

Figure 8.4 compares directly the performance of the Minimum Norm method, and the MUSIC method. For these simulations, the performance of the Minimum Norm method (and the MFBLP method based on Figure 8.2) was slightly better than MUSIC in terms of threshold performance. Above threshold there was no significant difference in the accuracy of any of these methods. Other researchers have also verified these results for two signal environments [8-3].

In comparing variants of MUSIC and the Minimum Norm methods, involving either the Johnson and Degraff approach, or the Wax and Kailath approach to the eigeninverse, no significant differences were found using the approach of Wax and Kailath and the whitened eigeninverse as shown in Figures 8.5(a) and (b). The approach of Johnson and Degraff was found to have poorer accuracy mainly due to the restriction to lower model orders (at higher model orders accuracy was significantly worse as indicated in Figure 8.1). The restriction to lower model orders, compared to the other approaches, also limits the maximum number of signal bearings that can be determined.

Using the QR factorization to compute an approximate whitened eigeninverse, there was no degradation in performance for the Minimum Norm method (shown in Figure 8.6 (a)), and only a slight degradation in threshold performance for the MUSIC method (shown in Figure 8.6 (b)).

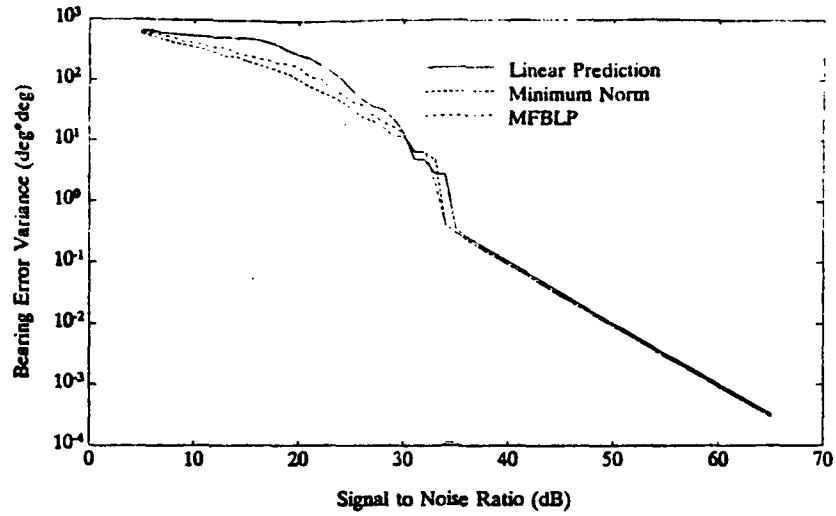


FIGURE 8.2: Comparison of the LP, MNorm, and MFBLP estimators

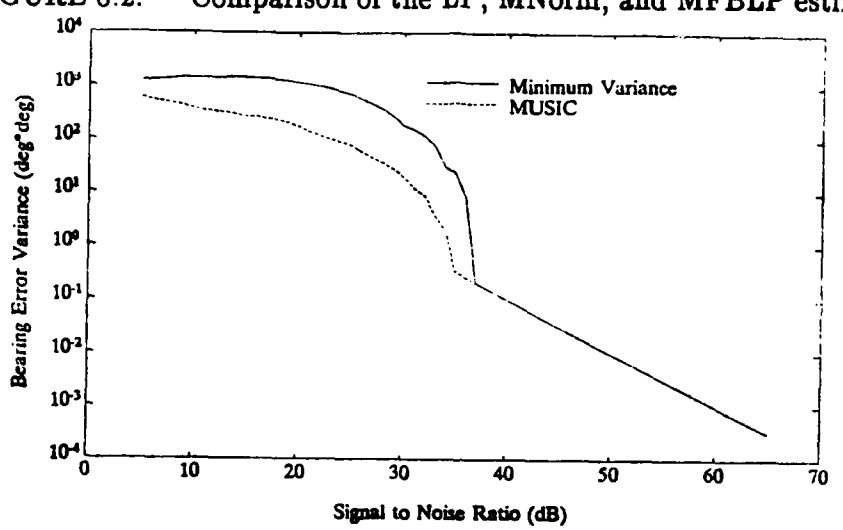


FIGURE 8.3: Comparison of the MV and MUSIC estimators

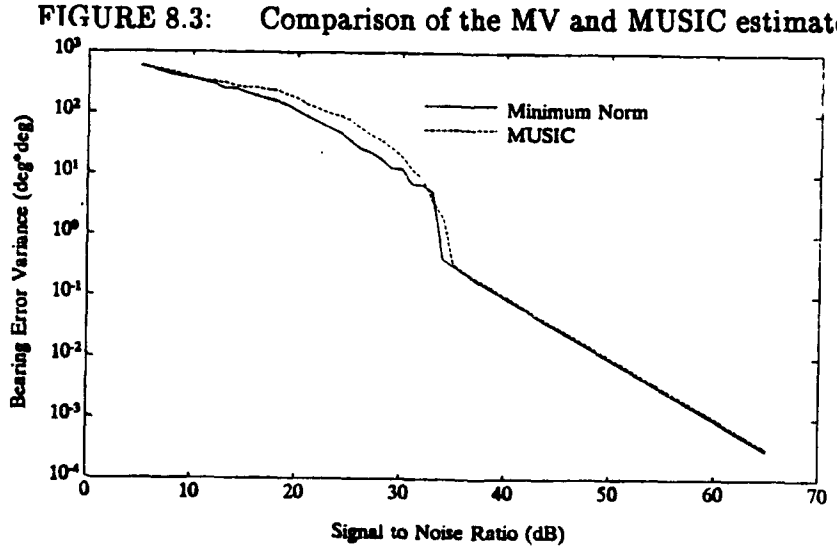
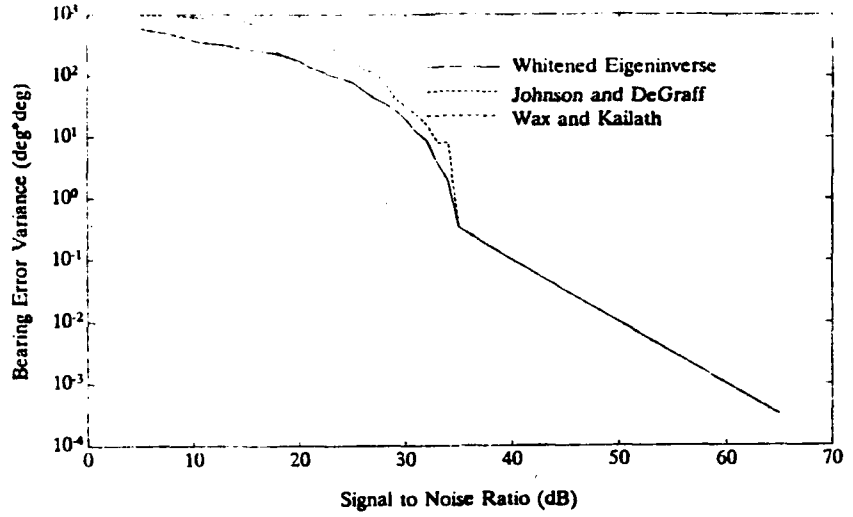
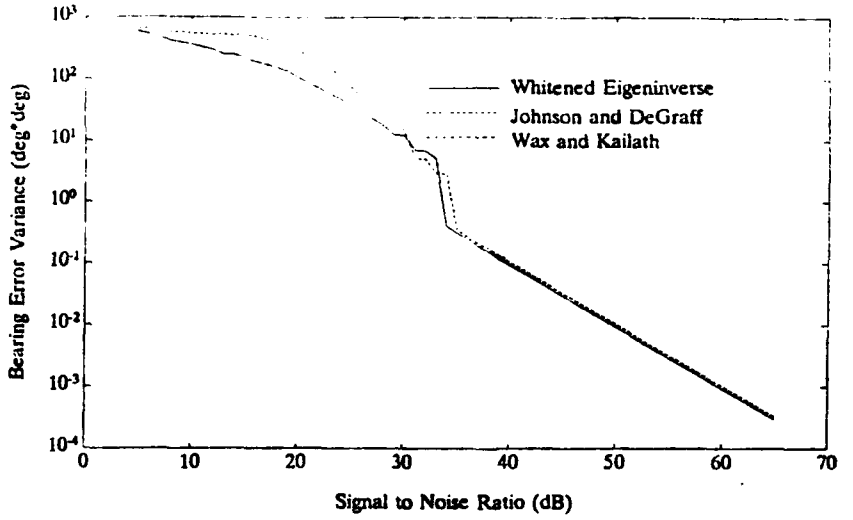


FIGURE 8.4: Comparison of the MUSIC and MNorm estimators

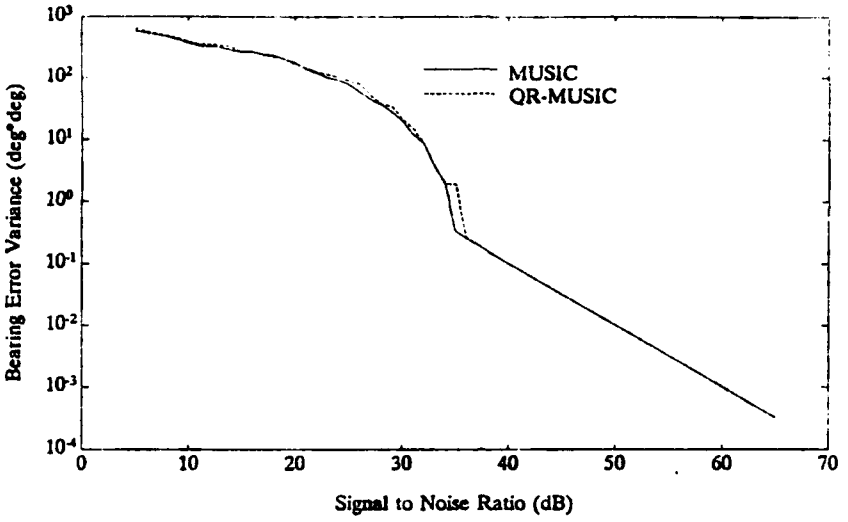


(b)

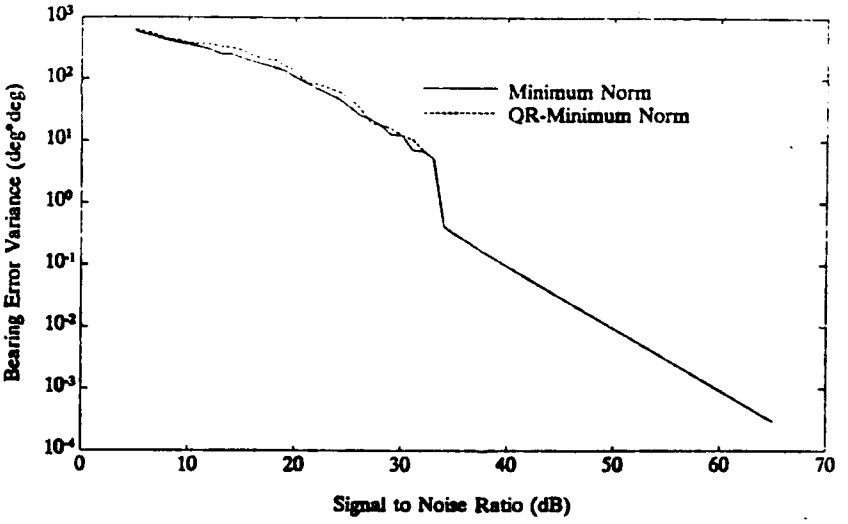


(a)

FIGURE 8.5: Comparison of three different eigeninverse approaches (a) Minimum Norm, and (b) MUSIC



(b)



(a)

FIGURE 8.6: Comparison of enhanced estimators with their QR versions (a) Minimum Norm, and (b) MUSIC

9.0 BEARING ESTIMATION USING THE ROOT METHOD

A problem with the method of searching the DF spectrum for signal peaks as outlined in Section 4.4 (called the spectral search method here) is that the steering vector e constrains the solutions to the form of a complex sinusoid. In the derivation of the DF estimators discussed in Section 5, no such constraint is placed on the computation of the coefficient vector **a**, the vector from which the DF spectrum is generated (see equation (4.29)). As a result, the solutions determined using the spectral method are not properly matched to the vector **a**. Figure 9.1 illustrates an example of this where a single peak indicates only one signal although two signals are present. A closer inspection of this peak, does reveal a "kink" which indicates the presence of the second signal but not an accurate estimate of its bearing.

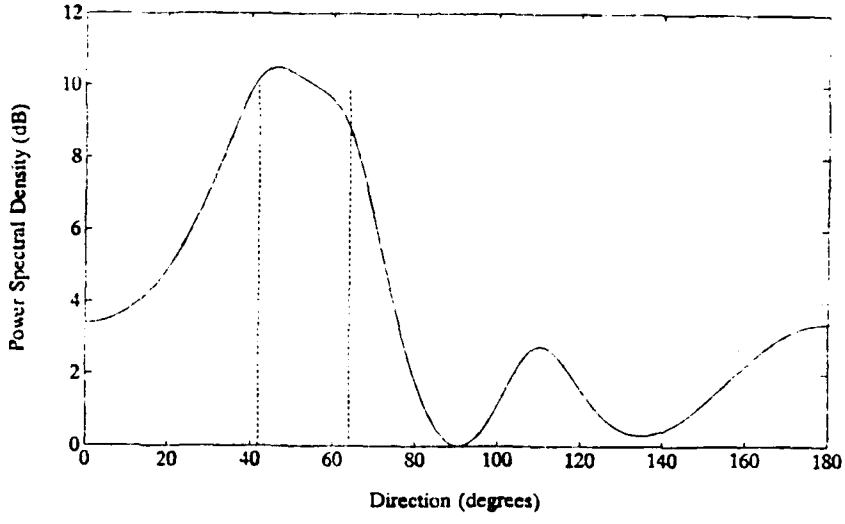


FIGURE 9.1: Minimum Variance bearing estimation using the DF spectrum (solid line) and the root method (dashed lines). The actual signals were at 40 and 60 degrees.

A better approach is to use a more general steering vector defined as,

$$\mathbf{s} = \frac{1}{\sqrt{p+1}} \begin{bmatrix} 1 \\ s \\ s^2 \\ \vdots \\ s^p \end{bmatrix}, \tag{9.1}$$

where

$$s = ce^{+j\omega d}, \tag{9.2}$$

in place of the steering vector $\mathbf{e}_p$ in equation (4.29) (see also equation (2.18)), that is,

$$S(\phi) = \frac{1}{\mathbf{s}^H\mathbf{a}\mathbf{a}^H\mathbf{s}}. \tag{9.3}$$

The difficulty here is that the search problem has been changed from a one dimensional problem in ω_s, to a two dimensional problem involving ω_s and c. This problem can be simplified by noting that maximizing equation (9.3) is equivalent to minimizing the term

$$\mathbf{s}^H\mathbf{a} = \sum_{n=0}^{p} a_n s^{-n}. \tag{9.4}$$

Since this is a polynomial equation, an equivalent mathematical representation is given by,

$$\mathbf{s}^H\mathbf{a} = a_0 \prod_{n=1}^{p}(1 - p_n s^{-1}), \tag{9.5}$$

where p_n represents the roots of this equation, but will be referred to as poles since p_n also represents half the poles of equation (9.3). (The other half of the poles of equation (9.3) are related by the expression,

$$p_{n+p} = -p_n^* \quad \text{for } 1 \leq n \leq p. \tag{9.6}$$

For the purposes of bearing estimation, however, determination of the poles given by equation (9.6) is unnecessary.)

Various rooting algorithms exist, although not discussed here, which are capable of determining the poles p_n for $1 \leq n \leq p$ given the coefficient vector $\mathbf{a}$. Once the poles have been determined, the corresponding bearings, based on equations (1.2) and (9.2), and also assuming the elevation angle $\psi = 0$, are given by

$$\phi = \cos^{-1}\left(\frac{\lambda}{2\pi d} \tan^{-1}\left(\frac{\mathrm{imag}\{p_k\}}{\mathrm{real}\{p_k\}}\right)\right). \tag{9.7}$$

In the previous example shown in Figure 9.1, the location of the two signals using the rooting method is shown. The improvement over the spectral method is obvious in this case. The improvement is also illustrated in the following simulation examples. The generation of the data is described in Section 5.4 and is also the same data as used in the examples in Section 8.3.

Figure 9.2 illustrates the improvement of the root-MUSIC and root-Minimum Norm methods compared to the spectral search version of the Minimum Norm method. The main improvement, which is true for any of the all pole methods, is better performance at lower signal to noise ratios (i.e. a lower threshold). Interestingly the improvement to the MUSIC method using the rooting method is greater than the improvement to the Minimum Norm method so that root-Music outperforms root-Minimum Norm. These results are consistent with observations made by other researchers for the two signal case [9-1]. Note that the error variance above the threshold is largely unaffected.

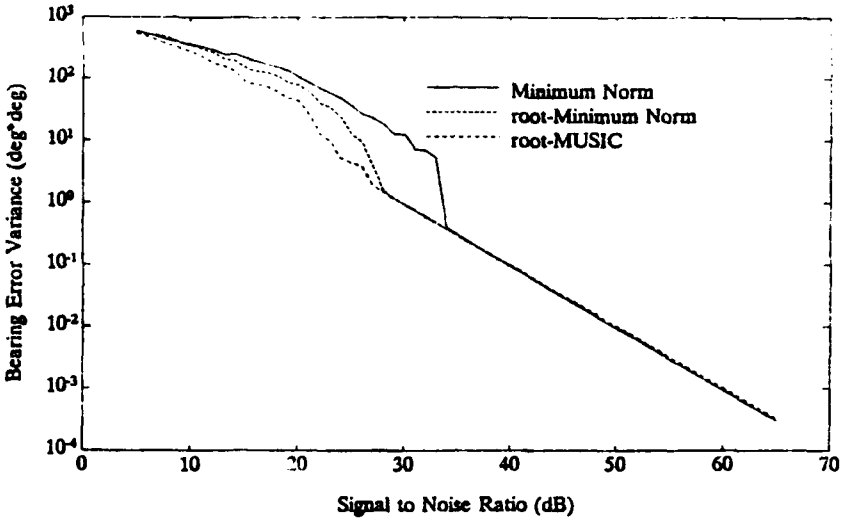
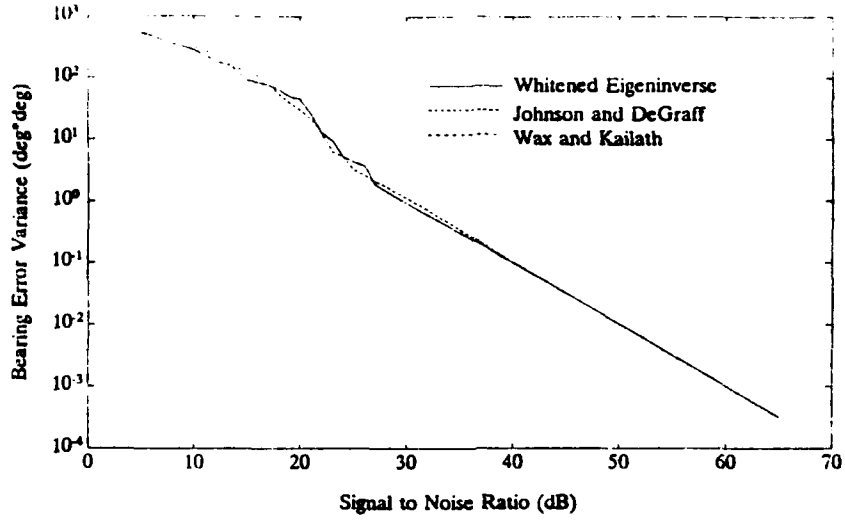
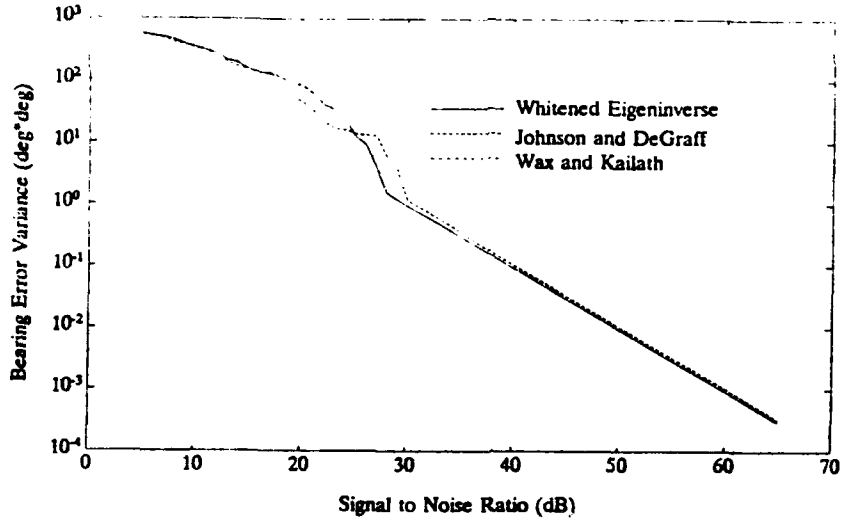


FIGURE 9.2: A comparison of two root methods and a spectral search method

Using variant approaches to calculate the eigeninverse, the qualitative differences were the same as the spectral search case. That is, there was no difference between root methods using the whitened eigeninverse or the approach of the Wax and Kailath, whereas the approach of Johnson and DeGraff resulted in poorer performance. Figures 9-3 (a) and (b) summarize these results.



(b)



(a)

FIGURE 9.3: A comparison of root based estimators using various eigeninverse approaches (a) root-Minimum Norm (b) root-MUSIC

10.0 DETERMINATION OF MODEL PARAMETERS

In all of the DF algorithms discussed so far, it has been assumed that the model order parameters such as the number of filter coefficients, p and q, and the number of signals, M, was already known. Generally this is not true, so that methods to estimate these parameters, which are discussed in the following sections, are required. However, before discussing these estimation methods, it is also useful to explore the fundamental limits on the maximum number of signals that a DF array can be used to exactly determine the bearing of.

10.1 FUNDAMENTAL LIMITS ON THE MAXIMUM NUMBER OF SIGNALS

Previously the signal environment, in terms of the n^{th} sensor input, was defined as,

$$x_n = \sum_{m=1}^{M} c_m(t) e^{-j(n\omega_m d + \theta_m)}. \tag{10.1}$$

In this analysis the effects of noise are ignored since the object is to determine the absolute maximum number of signal bearings which can be estimated as a function of the number of sensors.

At any instance in time, each signal can be completely described by three real valued parameters, namely, amplitude (represented by $|c_m(t)|$), phase (a function of both $c_m(t)$ and θ_m) and direction (or correspondingly the spatial frequency ω_m). Since these parameters may be arbitrarily chosen within a defined range, they are completely independent. In other words, if the signal environment is to be modelled, the signal model will require a minimum of $3M$ real valued parameters to completely describe the environment for a single measurement sample. For multiple samples the number of additional model parameters required is dependent on the type of signals (uncorrelated or correlated).

To begin with, the case involving signals which are all uncorrelated in time is examined. In this case, the values of θ_m and ω_m remain the same from sample to sample (since they are geometry dependent) but the values of the modulating envelope $c_m(t)$ (which is complex valued) will have changed. Assuming enough time has elapsed between samples so that consecutive values of $c_m(t)$ are independent, then each successive sample will increase the number of signal model parameters required by $2M$. If T samples are made then the total is given by,

$$N_p = 3M + 2M(T - 1). \tag{10.2}$$

In terms of the input data from which the model parameters are estimated, each sensor input x_n can be represented as a single complex value or by two real values, e.g. values representing amplitude and phase. For each measured sample, $2N$ real-valued input parameters are available for processing. The independence of these parameters, however, will be a function of the input signals, that is, the number of independent values cannot exceed the number generated by the signal model. Mathematically this can be stated as,

$$N_d = 2NT \qquad \text{for } N_p > 2NT, \tag{10.3}$$

otherwise,

$$N_d = N_p \qquad \text{for } N_p \leq 2NT, \tag{10.4}$$

where N_d is the number of independent real-valued input parameters available for processing; for brevity, these independent parameters will be referred to simply as the input parameters.

The condition $N_d = N_p$ is necessary for the solution of the signal model parameters. This condition does not necessarily guarantee a unique solution since ambiguities may exist depending on the signal model or array geometry (e.g. with linear arrays there is a 180 degree ambiguity). If $N_d < N_p$, however, no unique solution exists unless some form of a priori knowledge is incorporated into the solution (e.g. modulation type). Based on equations (10.2) and (10.3), this condition is met when,

$$2NT \geq 3M + 2M(T-1), \tag{10.5}$$

which in terms of M can be rewritten as,

$$M \leq \frac{2TN}{1+2T}. \tag{10.6}$$

For $T \geq \frac{N-1}{2}$, this simplifies to,

$$M \leq N-1. \tag{10.7}$$

The case involving signals which are all fully correlated in time can be approached by treating it as a single composite signal. This is motivated by the fact that not only do the values of θ_m and ω_m remain the same from sample to sample (since they are geometry dependent), but also the ratio of the modulating envelope $c_m(t)/c_1(t)$. Consequently only one single complex value (or two real values) is required to describe the change in the signal model parameters from sample to sample - analogous to a single signal model in the uncorrelated case.

For a single measurement sample, the composite signal will require $3M$ real-valued model parameters. Assuming enough time has elapsed between samples so that consecutive values of $c_m(t)$ are uncorrelated, then each successive sample will increase the number of real-valued signal model parameters required by 2. If T samples are made then the total is given by,

$$N_p = 3M + 2(T-1). \tag{10.8}$$

In terms of the input data, the analysis is similar to the uncorrelated case, with the exception that since the composite signal changes only by a complex multiplier term from sample to sample, each sensor sample will be related to the previous sample by the same constant multiplier term. In other words, after the first sample, only 2 input parameters are added to the input parameter set per sample. Based on this, the number of input parameters available after T samples is given by,

$$N_d = 2N + 2(T-1) \qquad \text{for } M > \frac{2N}{3} \tag{10.9}$$

and

$$N_d = 3M + 2(T - 1) \qquad \text{for } M \leq \frac{2N}{3}. \tag{10.10}$$

Comparing the first expression, equation (10.9), to equation (10.8) it is evident that for the stated conditions $N_d < N_p$ regardless of the number of samples. In comparing the second expression, equation (10.10), to equation (10.8), the two expressions are identical, that is, $N_d = N_p$. Clearly then, the upper limit on the number of correlated signals is given by,

$$M \leq \frac{2N}{3}. \tag{10.11}$$

Note this is identical to the uncorrelated case where only a single sample is used ($T = 1$).

Up to this point only the cases of uncorrelated signals only, and correlated signals only have been considered. The more general case consisting of a mixture of both types of signals can be solved in a similar manner as before by grouping the signals so that each group of signals consists of signals (called fundamental signals here) which are correlated with each other but are uncorrelated with signals from other groups. Each signal group is then treated as a single composite signal. In the following analysis the total number of fundamental signals (as before) is represented by M, the total number of composite signals or correlated groups is represented by K, and the total number of fundamental signals in each group is M_k.

Following the previous analysis for uncorrelated signals, the number of signal model parameters generated is given by,

$$N_p = 3M + 2K(T - 1), \tag{10.12}$$

where the first term and second terms in this expression are equivalent to the first and second terms respectively in equation (10.2).

The amount of input data available from the sensors will be as given in equations (10.3) and (10.4). Based on the previous analysis of correlated signals, the condition

$$M_k \leq \frac{2N}{3}, \tag{10.13}$$

is also imposed since adding more composite signals to the problem will not increase the upper limit on M_k.

The point at which the number of input parameters is sufficient, that is, $N_d = N_p$, occurs when

$$2NT \geq 3M + 2K(T - 1). \tag{10.14}$$

This expression can be rewritten in several ways to reveal some of the fundamental limits involved. For example, in terms of the number of composite signals K, this expression can be written as

$$K \leq \frac{2NT - 3M}{2(T-1)}, \tag{10.15}$$

which for $T \geq 1.5M - N + 1$ simplifies to,

$$K \leq N - 1. \tag{10.16}$$

The maximum limit on M, based on equation (10.14), can be written as,

$$M \leq \frac{2(N-K)T + 2K}{3}. \tag{10.17}$$

Inspection of this equation suggests no limit on M if the number of samples T is increased to infinity. However, realizing that

$$M = \sum_{k=1}^{K} M_k, \tag{10.18}$$

and letting $M_k = 2N/3$ (the upper limit for M_k) then an upper limit on M independent of the number of samples is given by,

$$M \leq \frac{2NK}{3}. \tag{10.19}$$

Equations (10.13), (10.15), and (10.17) to (10.19) are the general expressions for the mixed environment which place upper limits on the values of M_k, K, and M respectively. These limits reflect the fundamental limitations placed on the number of signals that may be estimated for any DF estimator. Whether a particular estimator can actually achieve these limits is dependent on the modelling approach involved. For example, in a signal environment where all signals are uncorrelated ($K = 1$) all the DF estimators discussed in this report are capable of estimating up to N-1 signal bearings (the upper limit) given a sufficient number of samples. In the fully correlated environment ($K = M$), DF estimators using the covariance method to estimate the autocorrelation matrix do not achieve the theoretical limit (only one half this value) while estimators using the modified covariance method do (see Section 3 for a discussion of these methods and their limitations). None of the estimators discussed in this report achieve the theoretical limits in the mixed signal environment ($1 < K < M$), and are at best limited to determining a maximum of N-1 signal bearings.

10.2 MODEL ORDER DETERMINATION

In the following discussion only the all pole methods discussed in Sections 5 and 8 are considered since MA methods have insufficient resolution for small tactical arrays, and ARMA methods lead to computational difficulties which are beyond the scope of this report. This limits the discussion to techniques which are used to determine the number of all pole filter coefficients, or filter order, p. Two types of signal environments are also considered: correlated and uncorrelated.

In the case of uncorrelated signals up to $N - 1$ signals can be processed, assuming a large number of sensor data samples (i.e. $T >> M$), without resorting to methods such as spatial smoothing. Since spatial smoothing results in a decrease in resolution, and tactical

arrays are generally small in terms of numbers of antenna, the most desirable choice for the filter order then is $p = N - 1$. The exception to this is for a small number of sensor data samples (i.e. $T < 10N$) where the autocorrelation matrix estimates may be unstable in which case the optimum filter order will lie somewhere between the maximum value of $N - 1$ and the optimum filter order for correlated signals discussed in the following paragraphs.

In the case of correlated signals, the choice of filter order is not nearly so straightforward. Spatial smoothing, which decreases the filter order, is required if all correlated sources are to be resolved. Decreasing the filter order also increases the stability of the autocorrelation matrix estimates since more data is involved in the computation of each element of the matrix. This can reduce the number of spurious estimates due to noise. However, decreasing the filter order also decreases the resolution of the estimator.

Figure 10.1 shows the effect of changing the filter order on the accuracy of the bearing estimates using the Linear Prediction, Minimum Variance, Thermal Noise, and Adaptive Angular Response methods. The plots represent the results of 100 trials of a 16 element (one third wavelength spacing) DF system used against 5 coherent signals with bearings of 40, 70, 100, 140, and 160 degrees. Bearing errors at the higher model orders ($7 < p < 14$) are a result of the appearance of spurious peaks in the spectrum.

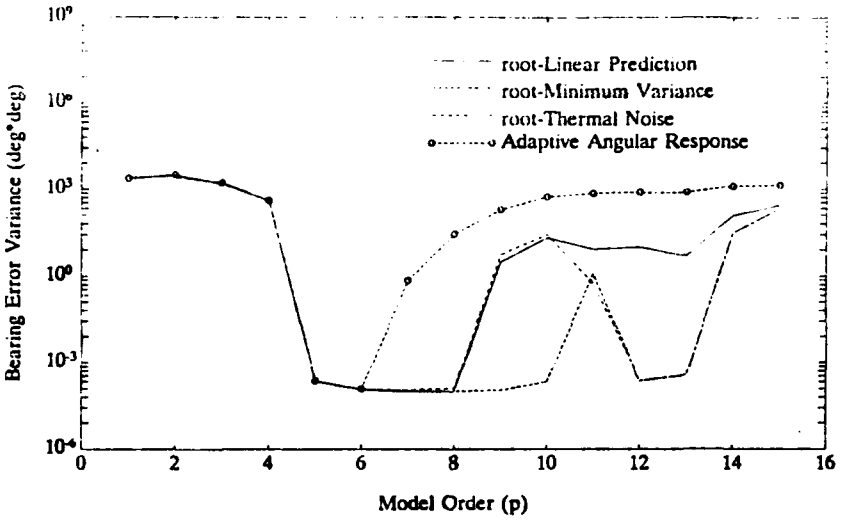


FIGURE 10.1: Bearing error variance as a function of the model order p for basic DF estimators

The optimum filter order for the Linear Prediction method has been shown analytically [10-1] to be about $p = N/3$ using either the covariance or modified covariance methods to estimate the autocorrelation matrix. Based on these analytical results and the simulation results shown in this report (Figures 5.1 and 10.1) a choice for the model order of $N/3 \leq p \leq N/2$ is appropriate. At higher filter orders, the extra coefficients give rise to spurious estimates in the spectrum which degrade the performance of the estimator. This choice also limits the number of signals that can be handled to $M < N/2$.

Given the similarity between the effect of model order on the accuracy of the Linear Prediction method and on the Thermal Noise method (see Figure 10.1), the choice for the model order is the same, namely, $N/3 \leq p \leq N/2$. The Minimum Variance method has greater immunity to spurious estimates so that slightly higher model orders are possible except that the accuracy of the method decreases at these higher orders so that a selection for the model order of $N/3 \leq p \leq N/2$ is again appropriate.

The Adaptive Angular Response method is the most sensitive to model order with the best choice being $p = M$ or slightly greater. Where the number of signals is unknown, algorithms, such as the ones discussed in Section 10.3 are required. This significantly increases the computational work load so that the advantage of using this method compared to the enhanced methods may be lost.

In the case of enhanced DF methods, taking advantage of the noise/signal subspace division effectively eliminates the spurious estimate problem. Figure 10.2 illustrates the improvement of the enhanced estimators using the same data as Figure 10.1. Researchers have determined that for these methods (using the whitened eigeninverse or the approach of Wax and Kailath) the optimum filter order lies somewhere between $0.6N$ and $0.8N$ [7-5],[10-2]. Based on maximizing the number of signal bearings which can be estimated, the best choice is $p = 2N/3$. For the methods using the approach of Johnson and DeGraff (e.g. the Eigenvector method), a lower model order is required due to a decrease in accuracy at higher model orders (see the example shown in Figure 10.3 at a model order of $p = 10$). This decrease in accuracy restricts the approach of Johnson and DeGraff to lower model orders, i.e. $p = N/2$.

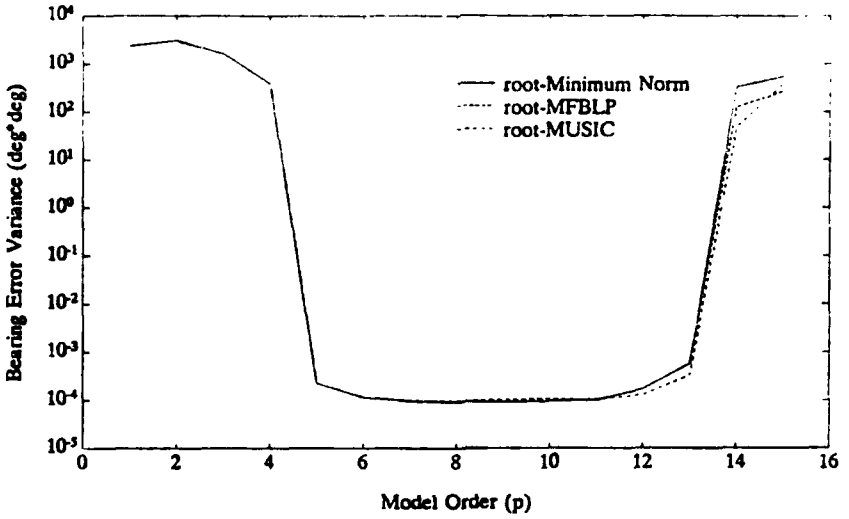


FIGURE 10.2: Bearing error variance as a function of the model order p for enhanced DF estimators

It should be noted that the optimum model orders described above have been considered for the case where only a single sensor sample is used for bearing estimation, and/or the signals are fully correlated. For a large number of samples ($T >> N$) and

uncorrelated signals, the optimum choice for the model order is $p = N - 1$. In land tactical VHF, correlated signals (multipath) are a fact of life, so that the previous analysis on model order selection is applicable.

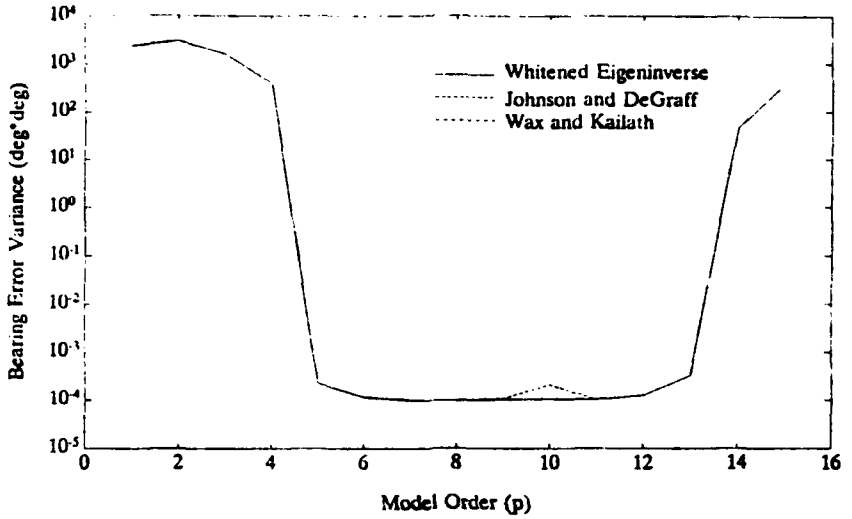


FIGURE 10.3: Bearing error variance as a function of the model order p for the root-MUSIC estimators using different eigeninverse approaches

10.3 SIGNAL NUMBER ESTIMATION

Enhanced all pole DF methods such as those discussed in Section 8 require knowledge of not only the filter order p, but also the number of signals M. Choosing M to be too small results in poor resolution and choosing M too large results in spurious estimates.

Kailath and Wax have reformulated the Akaike Information Criterion (AIC) and the Minimum Description Length (MDL) algorithms which were originally formulated for estimators using the autocorrelation method of autocorrelation matrix estimation [10-3],[10-4]. The reformulated versions use the eigenvalues of the normal (not augmented) autocorrelation matrix. The details of the derivation of these algorithms are found in reference [10-5], and the final results are given by,

$$\mathrm{AIC}(m) = N(p-m)\ln\left[\frac{\frac{1}{p-m}\sum_{i=m+1}^{p}\lambda_i}{\prod_{i=m+1}^{p}\lambda_i^{1/(p-m)}}\right] + m(2p-m), \qquad (10.20)$$

and,

$$\mathrm{MDL}(m) = N(p - m) \ln \left[\frac{\frac{1}{p-m} \sum_{i=m+1}^{p} \lambda_i}{\prod_{i=m+1}^{p} \lambda_i^{1/(p-m)}} \right] + \frac{1}{2} m(2p - m) \ln(N). \qquad (10.21)$$

The best estimate of the number of signals M is the value of the parameter m which minimizes either one of the above functions.

Studies have shown [10-6] that for a limited number of data samples the AIC algorithm performs best. However, this algorithm is not consistent, that is, as the number of samples is increased the probability of error does not decrease to zero. The MDL algorithm is consistent so that for large numbers of data samples it performs better than the AIC algorithm.

11.0 SUMMARY AND CONCLUSIONS

In this report, a number of the more popular superresolution DF methods were examined. In particular those methods which were based on adaptive filter models. Three filter models have evolved, namely, the moving average filter, the autoregressive filter, and the autoregressive moving average filter. In terms of accuracy and computational speed, methods based on the all pole or autoregressive filter model have shown the greatest promise and for this reason were the main focus of this report.

It was shown that the autoregressive filter based methods can be generalized as a five step procedure. These steps include:

1. Estimation of the autocorrelation matrix.

2. Division of the autocorrelation matrix into a signal and noise subspace.

3. Generation of an enhanced inverse autocorrelation matrix.

4. Estimation of the all pole filter coefficients.

5. Estimation of the signal bearings from the filter coefficients.

Each of these steps has been the focus of research on ways of optimizing the DF estimation process.

In terms of the resultant DF accuracy and computational simplicity, the modified covariance method has been found to be the best choice. Other methods including the autocorrelation and the covariance methods result in poorer accuracy, and maximum likelihood based methods, which potentially give higher accuracy, are currently too computationally intensive for realtime applications.

Enhancements to DF estimators based on the division of the autocorrelation matrix into a noise and signal subspace has proved to be extremely useful for a number of reasons. One is the suppression of spurious bearings estimates which can seriously degrade accuracy. A second reason is the ability to operate using higher order filters which leads to a slight improvement in accuracy and the ability to determine a greater number of signal bearings. A third reason is that methods which have been enhanced perform better at lower signal to noise ratios.

Two methods, eigen decomposition and singular value decomposition, are typically used to perform the signal/noise subspace division. A new autocorrelation matrix estimate is then generated using only the signal subspace. From a theoretical point of view, the results are identical using either decomposition method. An approximate method called the QR factorization method has also been proposed which achieves nearly the same performance as either eigen decomposition and singular value decomposition, but with significantly less processing requirements. This suggests that more research is required to determine how accurately the signal/noise subspace division must be done, and whether there are even faster methods that could be used.

The difficulty with using the signal subspace estimate for the autocorrelation estimate is that the resultant matrix is noninvertible. Since the inverse autocorrelation matrix is a central part of all the autoregressive filter based estimators, generation of an

enhanced inverse matrix is required. Four approaches have been proposed to do this, namely the pseudo-inverse, whitened eigeninverse, the approach by Johnson and Degraff, and the approach by Kailath and Wax. With the exception of the approach by Johnson and Degraff the differences between the various approaches were insignificant. The approach by Johnson and Degraff was found to be restricted to smaller model orders than the other approaches with a resultant degradation in accuracy and reduction in the number of signal bearings that can be estimated. From the standpoint of simplicity, the whitened eigeninverse is the preferred approach.

The methods for determining the autoregressive filter coefficients can be divided into two groups: Capon estimators (Minimum Variance and Thermal Noise) and linear prediction estimators (Autoregressive, Linear Prediction, and Maximum Entropy). Without the signal/noise subspace enhancements (steps 2 and 3) the linear prediction estimators are characterized by higher resolution abilities while the Capon estimators (at least in the case of the Minimum Variance estimator) are characterized by better suppression of spurious estimates. A third type of estimator, based on a pole-zero filter model, had the best threshold performance in the simulations performed for this report.

Of the enhanced estimators, the enhanced linear prediction estimators (e.g. Minimum Norm and MFBLP) have better threshold performance and correspondingly better resolution (when the bearings are determined from the DF spectrum) than the enhanced modified Capon estimators (i.e. MUSIC).

Two methods are currently in use for estimating the bearings based on the filter coefficients. The first and more commonly used method is based on a search of the DF spectrum which is generated based on the transfer function of the filter. This method suffers from a loss in resolution for two closely spaced signals (in bearing) at lower signal to noise ratios. A more recently popularized method is based on determining the bearings directly from the roots of the polynomial equation formed using the filter coefficients. The result is no merging of signal bearing as in the spectral search case, and a substantial improvement in performance at low signal to noise ratios. At high signal to noise ratios, the performance of both methods is identical. Based on this, the root method is clearly superior. Interestingly the root method brings the greatest improvement to the Capon estimators, that overall, the root-MUSIC method had the best performance.

Other considerations affecting DF include estimation of the number of signals and the optimum number of filter coefficients. Estimation of the number of signals is important for the enhanced methods since underestimating the number of signals leads to a serious degradation in accuracy and overestimating increases the possibility of spurious estimates which can also seriously degrade accuracy. Two algorithms, namely, Akaike Information Criteria (AIC) and Minimum Description Length (MDL), have been proposed. The AIC algorithm has better performance for a limited number of sensors and sensor data samples. For a larger number of samples the MDL algorithm performs better.

The choice of model order is important since this affects the accuracy of the resultant bearing estimates and the number of bearings that can be estimated. Additionally the basic DF methods are more susceptible to spurious estimates at higher model orders. For the basic methods, the optimum model order is generally the highest model order for which spurious estimates do not occur. Since this is a function of the number of signals which will generally be unknown (the AIC and MDL algorithms require eigendecomposition of the autocorrelation matrix and are therefore more appropriate for the enhanced methods) researchers have determined the best choice for the model order is between

$N/3 \leq p \leq N/2$ when the number of signals is small (i.e. $M \leq p$). For methods which are sensitive to model order, such as the Adaptive Angular Response, this criteria may not be sufficient to guarantee good results. For the enhanced methods researchers have found a choice of $0.6N \leq p \leq 0.8N$ to be appropriate.

Based on the results of comparisons between various methods in this report, the root-MUSIC estimator using the modified covariance autocorrelation matrix estimate represents the best approach. Approximate decomposition techniques such as QR factorization show great promise in decreasing the computational requirements of the DF estimation process without sacrificing accuracy, and this avenue of research should be explored more fully.

The concept of signal/noise subspace division has proved to be a major step forward in DF estimation. However the original concept is based on approximations which are valid for a large number of sensors. More research in the area of limited numbers of sensors is necessary.

It should be recognized that the comparisons in this report have focused primarily on artificial multipath type environments. The purpose of this is due to the fact that for tactical DF systems, the problems introduced by multipath in the operational environment are an order of magnitude more serious than any other error mechanism and remain largely unsolved. The word "artificial" is appropriate since there is very little documented results available on multipath measurements at VHF and UHF appropriate for DF research so that in this sense the multipath environments were contrived. It is quite likely that the numbers of multipath signals in the real environment, even for a single transmitter, will easily exceed the number of sensors of a tactical array. Under these conditions, the assumptions made in deriving the various DF methods will be in error and will require modifications. However, many of the techniques described in this report will still be applicable.

Two other somewhat artificial assumptions are the assumptions of white Gaussian noise statistics and perfect calibration of the array sensors. In practical systems these assumptions are rarely true and the result can be a severe degradation in the performance of a DF estimator. Although coloured noise can be handled by prewhitening the input sensor data using a preprocessor, this requires knowledge of the noise statistics beforehand (i.e. the noise only autocorrelation matrix). Determining unknown noise statistics and sensor calibration are both active areas of research.

It should also be recognized that the discussion in this report has been limited to linear arrays which have special mathematical advantages compared to nonuniform or nonlinear arrays. As a result some of the techniques described in this report are only applicable to uniform linear arrays. This includes spatial smoothing and rooting techniques. Additionally the modified covariance method is also restricted to symmetric arrays. A solution to this problem has been suggested in a paper by Friedlander [11-1] which describes how an arbitrary array can be interpolated to a linear array before processing. The result is that the techniques discussed in this report can be extended to arbitrary arrays, albeit with some extra computational complexity.

12.0 REFERENCES

1-1 Read, W.J.L., "Conventional VHF/UHF Land Tactical Direction Finding Systems", DREO TN 89-12, May 1989

2-1 Marple, S.L., *Digital Spectral Analysis with Applications*, Alan Oppenheim, editor, Prentice-Hall, Englewood Cliffs, New Jersey, Chapter 4, 1987.

2-2 Bartlett, M.S., "Smoothing Periodograms from Time Series with Continuous Spectra", *Nature*, vol. 161, pp. 686-687, May 1948.

2-3 Marple, S.L., *Digital Spectral Analysis with Applications*, Alan Oppenheim, editor, Prentice-Hall, Englewood Cliffs, New Jersey, Chapter 5, 1987.

2-4 Johnson, D.H., "The Application of Spectral Estimation Methods to Bearing Estimation Problems", *Proceedings of the IEEE*, vol. 70, no. 9, September 1982.

3-1 Makhoul, J., "Linear Prediction: A Tutorial Review", *Proceedings of the IEEE*, vol. 63., pp. 561-580, April 1975.

3-2 Shan, T., Wax, M., and Kailath, T., "On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 806-811, August 1985.

3-3 Marple, S.L., "A New Autoregressive Spectrum Analysis Algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 441-454, August 1980.

3-4 Pillai, S.U., and Kwon, B.H., "Forward/Backward Spatial Smoothing Techniques for Coherent Signal Identification", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 8-15, January 1989.

5-1 Cadzow, J.A., "Spectral Estimation: An Overdetermined Rational Model Equation Approach", *Proceedings of the IEEE*, vol. 70, no. 9, September 1982.

5-2 Burg, J.P., "Maximum Entropy Spectral Analysis", Ph.D. dissertation, Stanford University, Stanford, California, 1975.

5-3 Haykin, S., *Adaptive Filter Theory*, Prentice-Hall, Thomas Kailath, editor, Englewood Cliffs, New Jersey, Appendix B, 1986.

5-4 Marple, S., *Digital Spectral Analysis with Applications*, Alan Oppenheim, editor, Prentice-Hall, Englewood Cliffs, New Jersey, Chapter 7, 1987.

5-5 Ulrych, T.J., and Clayton, R.W., "Time Series Modeling and Maximum Entropy", *Phys. Earth Planet. Inter.*, vol 12, pp.188-200, August 1976.

5-6 Nutall, A.H., "Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Prediction Techniques", Naval Underwater Systems Center Technical Report TR-5303, New London, Conn., March 1976.

5-7 Capon, J., "High-resolution Frequency-Wavenumber Spectrum Analysis", *Proceedings of the IEEE,* vol. 57, pp. 1408-1419, August 1969.

5-8 Protter, M.H., and Morrey C.B., *Modern Mathematical Analysis,* Lynn Loomis, editor, Addison-Wesley Publishing Company, Reading, Mass., Chapter 4, 1964.

5-9 Gabriel, W.F., "Spectral Analysis and Adaptive Array Superresolution Techniques", *Proceedings of the IEEE,* vol. 68, no. 6, pp. 654-666, June 1980.

5-10 Borgiotti, G.V., and Kaplan, L.J., "Superresolution of Uncorrelated Interference Sources by Using Adaptive Array Techniques", *IEEE Transactions on Antennas and Propagation,* vol. 27, no. 6, pp. 842-845, November 1979.

5-11 Lagunas-Hernandez, M.,A., and Gasull-Llampallas, A., "An Improved Maximum Likelihood Method for Power Spectral Density Estimation", *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 32, no. 1, pp. 170-173, February 1984.

5-12 Welch, P.D., "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method based on Time Averaging over Short Modified Periodograms", *IEEE Transactions on Audio Electroacoustics,* vol. AU-15, pp. 70-73, June 1967.

6-1 Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vettering, W.T., *Numerical Recipes in C The Art of Scientific Computing,* Cambridge University Press, Cambridge, England, Chapter 11, 1988.

6-2 Eckart, C., and Young, G., "The Approximation of One Matrix by Another of Lower Rank", *Psychometrika,* vol. 1, no. 3, pp. 211-218, September 1936.

6-3 Klemma, V.C., and Laub, A.J., "The Singular Value Decomposition: Its Computation and Some Applications", *IEEE Transactions on Automatic Control,* vol. 25, pp. 164-176, April 1980.

6-4 Reilly, J.P, Chen. W.G., and Wong, K.M., "A Fast QR Based Array Processing Algorithm", *SPIE Proceedings of Advanced Algorithms and Architectures for Signal Processing III,* vol. 975, pp. 36-47, San Diego, California,15-17 August 1988.

7-1 Schmidt, R.O., "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Transactions on Antennas and Propagation,* vol. 34, no. 3, pp. 276-280, March 1986.

7-2 Albert, A., *Regression and the Moore–Penrose Pseudoinverse,* Academic Press, New York, 1972.

7-3 Johnson, D.H., and DeGraaf, S.R., "Improving the Resolution of Bearing in Passive Sonar Arrays by Eigenvalue Analysis", *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 30, no. 4, pp. 638-647, August 1982.

7-4 Wax, M., and Kailath, T., "Extending the Threshold of the Eigenstructure Methods", *International Conference on Acoustics, Speech and Signal Processing Proceedings,* ICASSP 85, pp. 556-559, 1985.

7-5 Tufts, D., and Kumaresan, R., "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood", *Proceedings of the IEEE*, vol. 70, no. 9, September 1982.

8-1 Kumaresan, R., and Tufts, D.W., "Estimating the Angles of Arrival of Multiple Plane Waves", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 19, no. 1, pp 134-138, January 1983.

8-2 Tufts, D.W., and Kumaresan, R., "Singular Value Decomposition and Improved Frequency Estimation Using Linear Prediction", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-30, pp.671-675, August 1982.

8-3 Kaveh, M., and Barabell, A.J., "The Statistical Performance of MUSIC and the Minimum-Norm Algorithms in Resolving Plane Waves in Noise", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 331-341, April 1986.

9-1 Rao, B.D., and Hari, K.V.S., "Statistical Performance Analysis of the Minimum-Norm Method", *IEE Proceedings*, vol. 136, Part F, no. 3, pp. 125-133, June 1989.

10-1 Lang, S.W., and McClellan, J.H., "Frequency Estimation with Maximum Entropy Spectral Estimators", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no.6, pp. 716-724, December 1980.

10-2 Clergeot, H., Tressens, S., and Ouamri, A., "Performance of High Resolution Frequencies Estimation Methods Compared to the Cramer-Rao Bounds", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no.11, pp. 1703-1720, November 1989.

10-3 Akaike, H., "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, vol AC-19, no. 6, pp. 716-723 December 1974.

10-4 Rissanen, J., "A Universal Prior for the Integers and Estimation by Minimum Description Length", *Ann. Stat.*, vol.11, pp. 417-431, 1983.

10-5 Wax, M., and Kailath, T., "Detection of Signals by Information Theoretic Criteria", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no.2, pp. 387-392, April 1985.

10-6 Zhang Q., Wong, K.M., Yip, P.C., and Reilly, J.P., "Statistical Analysis of the Performance of Information Theoretic Criteria in the Detection of the Number of Signals in Array Processing", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 10, pp. 1557-1567, October 1989.

11-1 Freidlander, B., "Direction Finding using an Interpolated Array", *International Conference on Acoustics, Speech and Signal Processing Proceedings*, ICASSP 90, pp. 2951-2954, Alburquerque, New Mexico, 1990.

13.0 GLOSSARY

13.1 MATHEMATICAL CONVENTIONS

$\mathbf{M}$	a boldface upper case letter represents a matrix
$\mathbf{v}$	a boldface lower case letter represents a column vector
s or S	a lower case or upper case italic letter represents a scalar quantity
*	complex conjugate
^{T}	matrix (or vector) transpose
^{H}	complex matrix (or vector) transpose
$\hat{}$	estimate
$^\Gamma$	matrix eigeninverse (defined in section 7.1.2)
$^\#$	matrix pseudoinverse (defined in section 7.1.2.1)
$E\{\}$	expectation of
real{}	real part of
imag{}	imaginary part of

13.2 SYMBOL DEFINITIONS

d	spacing between sensors
λ	signal wavelength
σ^2	noise variance
T	number of time samples
M	number of signals
N	number of sensors
ω	spatial frequency (as opposed to the usual definition for temporal frequency)
ϕ	bearing angle
ϕ_m	bearing angle of signal m
$x_m(t)$	complex baseband output of sensor m at time instance t
x_m	same as $x_m(t)$ except the time instance is unspecified
$\mathbf{x}$	sensor data vector (defined in section 3)
$\mathbf{X}$	sensor data matrix (defined in section 3)
$r_{xx}(m)$	autocorrelation lag m (defined in section 2.1.1)
$\mathbf{R}$	augmented autocorrelation matrix (defined in section 5.2.1)
$\mathbf{R}_p$	normal autocorrelation matrix (defined in section 5.2.1)
$\mathbf{R}_s$	autocorrelation matrix formed from signals only
$\mathbf{R}_n$	autocorrelation matrix formed from noise only
r_{ij}	element of the autocorrelation matrix occupying the i^{th} row and j^{th} column
σ_i	i^{th} singular value of the data matrix ordered so that $\sigma_i \geq \sigma_{i+1}$
λ_i	i^{th} eigenvalue of the autocorrelation matrix ordered so that $\lambda_i \geq \lambda_{i+1}$
$\mathbf{v}_i$	eigenvector corresponding to λ_i or right singular vector corresponding to σ_i
$\mathbf{u}_i$	left singular vector corresponding to σ_i
$\mathbf{e}$	steering vector (defined by equation 2.18)
$\mathbf{a}$	autoregressive filter coefficient vector
$\mathbf{b}$	moving average filter coefficient vector
p	autoregressive (all pole) filter order - p+1 is also the spatial smoothing subarray size

q	moving average (all zero) filter order
$S(\phi)$	direction finding spectrum

13.3 ACRONYMS

AAR	adaptive angular response
AIC	Akaike information criterion
AR	autoregressive
ARMA	autoregressive moving average
BART	Bartlett
DF	direction finding
EV	eigenvector
MDL	minimum description length
MUSIC	multiple signal classification
MV	minimum variance
TN	thermal noise

APPENDIX A - POLYNOMIAL FACTORIZATION

In classical and superresolution DF estimators functions of the form

$$G(\omega) = \mathbf{e}^{H}\mathbf{D}\mathbf{e}, \tag{A1}$$

are often encountered where e is a q+1 element steering vector of the form,

$$\mathbf{e} = [\, 1,\, e^{-j\omega d},\, e^{-j\omega 2d},\, \ldots,\, e^{-j\omega q d}]^{T} \tag{A2}$$

and D is a $(q+1) \times (q+1)$ Hermitian symmetric matrix (such as the autocorrelation matrix or its inverse). In expanded form, the function $G(\omega)$ may also be represented as,

$$G(\omega) = \sum_{n=-q}^{q} a_n e^{-j\omega n d}. \tag{A3}$$

where

$$a_n = \sum_{i=0}^{q-n} d_{i,i+n}, \tag{A4}$$

and $d_{i,j+n}$ is the element located in the i^{th} row and $j+n^{th}$ column of the matrix D. Given the properties of the matrix D, the coefficients a_n will also have the property that,

$$a_n = a_{-n}. \tag{A5}$$

Since equation (A3) represents a polynomial expression with an odd number of coefficients, it can be decomposed into the product of simpler polynomial terms given by,

$$G(\omega) = r_{zz}(0) \prod_{n=1}^{q} (-c_n^{*} e^{+j\omega d} + 1 - c_n e^{-j\omega d}). \tag{A6}$$

The polynomial term on the right hand side of this expression can be expressed in terms of its roots as,

$$-c_n^{*} e^{+j\omega d} + 1 - c_n e^{-j\omega d} = (1 - p_{1n} e^{-j\omega d})(1 - p_{2n} e^{-j\omega d}) \frac{-e^{+j\omega d}}{p_{1n} + p_{2n}}. \tag{A7}$$

From inspection of equation (A7), if p_n is a root, then $1/p_n^{*}$ must also be a root. This fact can be used to advantage although two separate cases must be considered. In the first case, where $|p_n| \neq 1$, the roots exist in pairs which must satisfy the inverse conjugate relationship. Under these conditions the coefficients c_n are chosen so that

$$p_{1n} = p_n \tag{A8}$$

and

$$p_{2n} = 1/p_n^* \tag{A9}$$

Using these new relationships equation (A7) becomes,

$$-c_n^* e^{+j\omega d} + 1 - c_n e^{-j\omega d} = (1 - p_n e^{-j\omega d})(1 - \frac{e^{-j\omega d}}{p_n^*}) \frac{-p_n^* e^{+j\omega d}}{p_n p_n^* + 1}$$

$$= (1 - p_n e^{-j\omega d})(1 - p_n^* e^{+j\omega d}) \frac{1}{p_n p_n^* + 1}. \tag{A10}$$

In the second case where $|p_n| = 1$, the root pairs do not necessarily exist in pairs since each root of unit magnitude is its own conjugate inverse, that is,

$$p_n = \frac{1}{p_n^*}. \tag{A11}$$

However, assuming that the function $G(\omega)$ has the property that

$$G(\omega) \geq 0 \quad \text{for all } \omega, \tag{A12}$$

(such as when $G(\omega)$ represents the true spectrum or its inverse) and ignoring cases where different quadratic polynomial terms have the same roots (in that case the c_n are chosen so that $p_{1n} = p_{2n}$ so the analysis used when $|p_n| \neq 1$ still applies) then it also follows that

$$-c_n^* e^{+j\omega d} + 1 - c_n e^{-j\omega d} \geq 0, \tag{A13}$$

By inspection (e.g. letting $\omega_s = 0$),

$$c_n \leq \frac{1}{2}. \tag{A14}$$

Equation (A7) is a quadratic polynomial so that the two poles p_{1n} and p_{2n} can be determined from c_n using the quadratic equation. This gives,

$$p_n = \frac{1 \pm \sqrt{1 - 4c_n c_n^*}}{2c_n}. \tag{A15}$$

Since $|p_n| = 1$, then

$$1 = \left| \frac{1 \pm \sqrt{1 - 4c_n c_n^*}}{2c_n} \right|. \tag{A16}$$

Rearranging this result in terms of c_n and using the fact that $|c_n| \leq 1/2$, then

$$\sqrt{4c_n c_n^*} = 1 \pm \sqrt{1 - 4c_n c_n^*}. \tag{A17}$$

Squaring both sides,

$$4c_n c_n^* = 1 \pm 2\sqrt{1 - 4c_n c_n^*} + (1 - 4c_n^* c_n), \tag{A18}$$

and simplifying,

$$-(1 - 4c_n c_n^*) = \pm\sqrt{1 - 4c_n c_n^*}. \tag{A19}$$

Again squaring both sides

$$(1 - 4c_n c_n^*)^2 = 1 - 4c_n c_n^*, \tag{A20}$$

and simplifying

$$4c_n c_n^*(1 - 4c_n c_n^*) = 0. \tag{A21}$$

The solution $c_n c_n^* = 0$ is not a valid solution to equation (A15) so

$$c_n c_n^* = \frac{1}{4}. \tag{A22}$$

Based on this last result, equation (A15) simplifies to,

$$p_n = \frac{1}{2c_n} \tag{A23}$$

Since there is only one solution for both p_{1n} and p_{2n}, they both must be equal. That is,

$$p_{1n} = p_{2n} = p_n \tag{A24}$$

These roots also satisfy the relationships expressed in equations (A8) and (A9), and as a result equation (A10) is still valid for this case.

From the preceding analysis, if the condition $G(\omega) \geq 0$ is met for all ω, and using the result given in equation (A10), then equation (A6) can be factored into two conjugate parts as,

$$G(\omega) = r_{zz}(0) \prod_{n=1}^{q} \frac{(1 - p_n e^{-j\omega d})}{\sqrt{p_n p_n^* + 1}} \prod_{n=1}^{q} \frac{(1 - p_n e^{+j\omega d})}{\sqrt{p_n p_n^* + 1}}. \tag{A25}$$

Finally, converting each multiplier term back to polynomial form, the result is

$$G(\omega) = \sum_{n=0}^{q} b_n e^{-j\omega nd} \sum_{n=0}^{q} b_n^* e^{+j\omega nd} \tag{A26}$$

where the coefficients represented by b_n are a result of combining terms with the same power of e.

APPENDIX B - OPTIMUM REDUCED RANK MATRICES

The problem to be examined in this discussion is that given an $m \times n$ matrix $\mathbf{A}$ of rank K, what is the best reduced rank (less than K) matrix equivalent of $\mathbf{A}$? The definition of best reduced rank matrix can be given a mathematical formulation by considering the error matrix,

$$\mathbf{E} = \mathbf{A} - \mathbf{B}, \tag{B1}$$

where $\mathbf{B}$ is an $m \times n$ reduced rank equivalent of matrix $\mathbf{A}$. The best matrix $\mathbf{B}$, in the least squares sense minimizes the error power given by,

$$\epsilon^2 = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} e_{ij} e_{ij}^{*}, \tag{B2}$$

where e_{ij} represents an element of the matrix $\mathbf{E}$ in equation (B1).

One approach to solving this problem is to consider an orthonormal vector basis set, $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{K-1}$ (where $\mathbf{v}_i$ is an n element column vector), used to represent the rows of the matrix $\mathbf{A}$. For example, if $\mathbf{a}_i^H$ is an n element row vector representing the i^{th} row of the matrix, then

$$\mathbf{a}_i^H = \sum_{k=0}^{K-1} \gamma_{ik} \mathbf{v}_k^H. \tag{B3}$$

Multiplying both sides of this expression by $\mathbf{v}_j$ then

$$\begin{aligned} \mathbf{a}_0^H \mathbf{v}_j &= \gamma_{0j}, \\ \mathbf{a}_1^H \mathbf{v}_j &= \gamma_{1j}, \\ \mathbf{a}_2^H \mathbf{v}_j &= \gamma_{2j}, \\ &\vdots \\ \mathbf{a}_{m-1}^H \mathbf{v}_j &= \gamma_{m-1,j}, \end{aligned} \tag{B4}$$

since by definition $\mathbf{v}_i^H \mathbf{v}_j = 1$ if $i = j$ and $\mathbf{v}_i^H \mathbf{v}_j = 0$ if $i \neq j$.

Based on the set of equations (B4) a second vector set can be defined such that,

$$\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \tag{B5}$$

where $\mathbf{u}_i$ is an m element normalized column vector, and σ_j is a positive real scalar. The matrix $\mathbf{A}$ can now be decomposed in terms of the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ as

$$\mathbf{A} = \sum_{j=0}^{K-1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H. \tag{B6}$$

Using this expression, a reduced rank matrix $\mathbf{B}$ can be formed simply by setting the appropriate number of scalar coefficients σ_j to zero.

To simplify the problem, it is assumed that the values of σ_j are ordered so that for a rank reduction of 1, $\sigma_{K-1} = 0$, for a rank reduction of 2, $\sigma_{K-1} = \sigma_{K-2} = 0$, etc. Examining the case where the rank is to be reduced by 1, the reduced rank matrix may be described by,

$$\mathbf{B} = \sum_{j=0}^{K-2} \sigma_j \mathbf{u}_j \mathbf{v}_j^H. \tag{B7}$$

Using equation (B6) for the definition of $\mathbf{A}$, and equation (B7) for the definition of $\mathbf{B}$, then the error matrix defined by equation (B1) becomes

$$\mathbf{E} = \sum_{j=0}^{K-1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H - \sum_{j=0}^{K-2} \sigma_j \mathbf{u}_j \mathbf{v}_j^H. \tag{B8}$$

and simplifying

$$\mathbf{E} = \sigma_{K-1} \mathbf{u}_{K-1} \mathbf{u}_{K-1}^H \tag{B9}$$

If u_i is an element of the vector $\mathbf{u}_{K-1}$ and v_i is an element of the vector $\mathbf{v}_{K-1}$, then the elements of the matrix $\mathbf{E}$ are given by,

$$e_{ij} = \sigma_{K-1} u_i \overset{*}{v_j}$$

and the error power is given by,

$$\epsilon^2 = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (\sigma_{K-1} \overset{*}{u_i} v_j)(\sigma_{K-1} u_i \overset{*}{v_j}). \tag{B10}$$

This expression can be simplified by first rearranging the coefficients,

$$\epsilon^2 = \sigma_{K-1}^2 \sum_{i=1}^{m} \overset{*}{u_i} u_i \sum_{j=1}^{n} v_j \overset{*}{v_j}. \tag{B11}$$

and then recalling that $\mathbf{v}_{K-1}^H \mathbf{v}_{K-1} = \mathbf{u}_{K-1}^H \mathbf{u}_{K-1} = 1$, therefore

$$\epsilon^2 = \sigma_{K-1}^2. \tag{B12}$$

Using this last result, the error power will be minimized if σ_{K-1}^2 is minimized. In terms of the choice of the basis vector $\mathbf{v}_{K-1}$, and utilizing the relationship in equation (B5), equation (B12) rewritten as,

$$\epsilon^2 = \sigma_{K-1}^2 = (\sigma_{K-1} \mathbf{u}_{k-1})^H (\sigma_{K-1} \mathbf{u}_{k-1}) = \mathbf{v}_{K-1}^H \mathbf{A}^H \mathbf{A} \mathbf{v}_{K-1}. \tag{B13}$$

Inspection of equation (B13) suggests an alternate description of the problem. That is, choose a direction $\mathbf{v}_{K-1}$ in the vector space defined by the rows of $\mathbf{A}$ (other than the null vector) for which the sum of the power of the row vector components (γ_{iK-1} in equation (B4)) in that direction is minimized.

Minimizing ϵ^2 with respect to each of the elements of $\mathbf{v}_{K-1}$ with the constraint that $\mathbf{v}_{K-1}^H\mathbf{v}_{K-1} = 1$ is simplified using the Lagrange multiplier technique [5-8]. Defining a new function,

$$F(\mathbf{v}_{K-1}) = \mathbf{v}_{K-1}^H\mathbf{A}^H\mathbf{A}\mathbf{v}_{K-1} + \lambda_{K-1}(1 - \mathbf{v}_{K-1}^H\mathbf{v}_{K-1}), \tag{B14}$$

this incorporates both equation (B13) and the unit length constraint for $\mathbf{v}_{K-1}$. The solution is derived by minimizing this equation with respect to each of the elements of $\mathbf{v}_{K-1}$. The result is,

$$2\mathbf{A}^H\mathbf{A}\mathbf{v}_{K-1} - 2\lambda_{K-1}\mathbf{v}_{K-1} = 0. \tag{B15}$$

Rearranging gives,

$$\mathbf{A}^H\mathbf{A}\mathbf{v}_{K-1} = \lambda_{K-1}\mathbf{v}_{K-1}, \tag{B16}$$

which is an eigenvector expression for $\mathbf{A}^H\mathbf{A}$.

Returning once again to the error power expression, equation (B13), and incorporating the latest result, then

$$\epsilon^2_{\min} = \mathbf{v}_{K-1}^H\mathbf{A}^H\mathbf{A}\mathbf{v}_{K-1} = \mathbf{v}_{K-1}^H\lambda_{K-1}\mathbf{v}_{K-1} = \lambda_{K-1}. \tag{B17}$$

Therefore the minimum error power $\epsilon^2_{\min}$ results if the eigenvector associated with the smallest eigenvalue of the matrix $\mathbf{A}^H\mathbf{A}$ is used for the basis vector $\mathbf{v}_{K-1}$.

Extending this analysis to the case where the rank of matrix $\mathbf{A}$ is reduced by any number, it is apparent that the eigenvectors of the matrix $\mathbf{A}^H\mathbf{A}$ should be ordered so that the corresponding eigenvalues are arranged in decreasing order, that is, $\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{K-1}$. The optimum reduced rank matrix can then be computed by using equation (B7) and setting the limit of the summation equal to the rank of the reduced matrix.

From the preceding analysis it is apparent that the basis vectors, $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{K-1}$, can be computed by performing an eigenanalysis of the matrix $\mathbf{A}^H\mathbf{A}$. The corresponding vectors $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{K-1}$, and scalar values $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{K-1}$ can then be computed using the relationship expressed by equation (B5). Alternatively, noting vector set $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ is orthonormal (proof: Equation (B5) can be modified as,

$$
\begin{aligned}
(\sigma_i \mathbf{u}_i)^H(\sigma_j \mathbf{u}_j) &= (\mathbf{A}\mathbf{v}_i)^H(\mathbf{A}\mathbf{v}_j) \\
\sigma_i\sigma_j \mathbf{u}_i^H \mathbf{u}_j &= \mathbf{v}_i^H(\mathbf{A}^H\mathbf{A}\mathbf{v}_j) \\
\sigma_i\sigma_j \mathbf{u}_i^H \mathbf{u}_j &= \mathbf{v}_i^H(\lambda_j \mathbf{v}_j) \\
\mathbf{u}_i^H \mathbf{u}_j &= \frac{\lambda_j}{\sigma_i\sigma_j}\, \mathbf{v}_i^H \mathbf{v}_j.
\end{aligned} \tag{B19}
$$

If $i \neq j$ then

$$
\mathbf{u}_i^H \mathbf{u}_j = \frac{\lambda_j}{\sigma_i\sigma_j}\, \mathbf{v}_i^H \mathbf{v}_j = \frac{\lambda_j}{\sigma_i\sigma_j}(0) = 0, \tag{B20}
$$

and if $i = j$ then

$$
\mathbf{u}_i^H \mathbf{u}_i = \frac{\lambda_i}{\sigma_i^2}\, \mathbf{v}_i^H \mathbf{v}_i = \frac{\lambda_i}{\lambda_i}(1) = 1, \tag{B21}
$$

where the relationship $\lambda_i = \sigma_i^2$ follows from equations (B12) and (B17).), and modifying equation (B5) in the following manner:

$$
\begin{aligned}
(\mathbf{A}\mathbf{v}_j)^H &= (\sigma_j \mathbf{u}_j)^H \\
\mathbf{v}_j^H \mathbf{A}^H &= \sigma_j \mathbf{u}_j^H \\
\mathbf{v}_j^H \mathbf{A}^H \mathbf{u}_j &= \sigma_j (\mathbf{v}_j^H \mathbf{v}_j)\mathbf{u}_j^H \mathbf{u}_j \\
\mathbf{A}^H \mathbf{u}_j &= \sigma_j \mathbf{v}_j
\end{aligned} \tag{B22}
$$

By symmetry then, the vector set $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{K-1}$ is an orthonormal basis vector set for the columns of matrix $\mathbf{A}$ [6-3]. Following the same analysis as previously then $\mathbf{u}_j$ can be shown to be an eigenvector of the matrix $\mathbf{A}\mathbf{A}^H$ with the same non-zero eigenvalue λ_j as the matrix $\mathbf{A}^H\mathbf{A}$.

Given that both vector sets $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{K-1}$ and $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{K-1}$ are orthonormal vector sets, and the values $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{K-1}$ are positive real numbers, then equation (B6) defines the singular value decomposition of matrix $\mathbf{A}$. In this case the vectors $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{K-1}$ are called the left singular vectors, the vectors $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{K-1}$ are called the right singular vectors, and the positive real scalar values $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_{K-1}$ are called the singular values. Using this fact, the singular vectors and singular values may be calculated directly using singular value decomposition techniques. Again for the purposes of optimum reduced rank matrices, the ordering of the vectors will be such that for the corresponding singular values $\sigma_0 \geq \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{K-1}$ (since $\sigma_j^2 = \lambda_j$ as established by equations (B12) and (B17)).

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

NATIONAL DEFENCE
DEFENCE RESEARCH ESTABLISHMENT OTTAWA
SHIRLEY BAY, OTTAWA, ONTARIO K1A 0K2 CANADA

2. SECURITY CLASSIFICATION (overall security classification of the document. including special warning terms if applicable)

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

AN EVALUATION OF SUPERRESOLUTION METHODS FOR TACTICAL RADIO DIRECTION FINDING (U)

4. AUTHORS (Last name, first name, middle initial)

READ, WILLIAM J.L.

5. DATE OF PUBLICATION (month and year of publication of document)

OCTOBER 1991

6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)

121

6b. NO. OF REFS (total cited in document)

41

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

TECHNICAL REPORT

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

NATIONAL DEFENCE
DEFENCE RESEARCH ESTABLISHMENT OTTAWA
SHIRLEY BAY, OTTAWA, ONTARIO K1A 0K2 CANADA

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)

041LK11

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DREO REPORT 1091

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availabilty (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

(U) This report evaluates superresolution direction finding methods for land tactical narrowband VHF/UHF purposes. The DF methods are described and evaluated in depth using a common theoretical framework beginning with classical methods and ending with modern day eigenanalysis based methods. Based on this analysis, superresolution methods can be described in terms of a five step procedure where each step has a unique purpose in the estimation process. Different methods vary according to the operation performed at each step. This is useful in analyzing and comparing the theoretical performance of various DF methods under the same conditions. The results from simulations are also included to support the theoretical evaluations.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military pr ect code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

SUPERRESOLUTION
DIRECTION FINDING
NARROWBAND
RADIO
EIGENANALYSIS
SINGULAR VALUE DECOMPOSITION
AUTOREGRESSIVE
LINEAR PREDICTION
MINIMUM VARIANCE
MUSIC
MINIMUM NORM
SPECTRAL ESTIMATION

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM